

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/64, 15/70		A1	(11) International Publication Number: WO 92/06204 (43) International Publication Date: 16 April 1992 (16.04.92)
<p>(21) International Application Number: PCT/US91/07149</p> <p>(22) International Filing Date: 27 September 1991 (27.09.91)</p> <p>(30) Priority data: 590,219 28 September 1990 (28.09.90) US</p> <p>(71) Applicant: IXSYS, INC. [US/US]; 3550 General Atomics Court, Suite L103, San Diego, CA 92121 (US).</p> <p>(72) Inventor: HUSE, William, D. ; 471 Avenida Primavera, Del Mar, CA 92014 (US).</p> <p>(74) Agents: CAMPBELL, Cathryn et al.; Pretty, Schroeder, Brueggemann & Clark, 444 South Flower Street, Suite 2000, Los Angeles, CA 90071 (US).</p>			
<p>(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC, MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PL, RO, SD, SE (European patent), SN (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent).</p>			

Published

With international search report.

General →

(54) Title: SURFACE EXPRESSION LIBRARIES OF HETEROMERIC RECEPTORS

(57) Abstract

A composition of matter comprising a plurality of prokaryotic cells containing diverse combinations of first and second DNA sequences encoding first and second polypeptides which form a heteromeric receptor exhibiting binding activity toward a preselected molecule, said heteromeric receptors being expressed on the surface of filamentous bacteriophage.

+ DESIGNATIONS OF "SU"

Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	CN	Guinea	NL	Netherlands
BJ	Benin	CR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	R	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU*	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TC	Togo
DE	Germany	US	United States of America	US	United States of America
DK	De-				

SURFACE EXPRESSION LIBRARIES
OF HETEROGENERIC RECEPTORS

BACKGROUND OF THE INVENTION

This invention relates generally to recombinant
5 expression of heteromeric receptors and, more particularly,
to expression of such receptors on the surface of
filamentous bacteriophage.

Antibodies are heteromeric receptors generated by a
vertebrates organism's immune system which bind to an
10 antigen. The molecules are composed of two heavy and two
light chains disulfide bonded together. Antibodies have
the appearance of a "Y" - shaped structure and the antigen
binding portion being located at the end of both short arms
of the Y. The region on the heavy and light chain
15 polypeptides which corresponds to the antigen binding
portion is known as variable region. The differences
between antibodies within this region are primarily
responsible for the variation in binding specificities
between antibody molecules. The binding specificities are
20 a composite of the antigen interactions with both heavy and
light chain polypeptides.

The immune system has the capability of generating an
almost infinite number of different antibodies. Such a
large diversity is generated primarily through
25 recombination to form the variable regions of each chain
and through differential pairing of heavy and light chains.
The ability to mimic the natural immune system and generate
antibodies that bind to any desired molecule is valuable
because such antibodies can be used for diagnostic and
30 therapeutic purposes.

Until recently, generation of antibodies against a

desired molecule was accomplished only through manipulation of natural immune responses. Methods included classical immunization techniques of laboratory animals and monoclonal antibody production. Generation of monoclonal 5 antibodies is laborious and time consuming. It involves a series of different techniques and is only performed on animal cells. Animal cells have relatively long generation times and require extra precautions to be taken compared to procaryotic cells to ensure viability of the cultures.

10 A method for the generation of a large repertoire of diverse antibody molecules in bacteria has been described, Huse et al., Science, 246, 1275-1281 (1989), which is herein incorporated by reference. The method uses the bacteriophage lambda as the vector. The lambda vector is 15 a long, linear double-stranded DNA molecule. Production of antibodies using this vector involves the cloning of heavy and light chain populations of DNA sequences into separate vectors. The vectors are subsequently combined randomly to form a single vector which directs the coexpression of 20 heavy and light chains to form antibody fragments. A disadvantage to this method is that undesired combinations of vector portions are brought together when generating the coexpression vector. Although these undesired combinations do not produce viable phage, they do however, result in a 25 significant loss of sequences from the population and, therefore, a loss in diversity of the number of different combinations which can be obtained between heavy and light chains. Additionally, the size of the lambda phage gene is large compared to the genes that encode the antibody 30 segments. This makes the lambda system inherently more difficult to manipulate as compared to other available vector systems.

There thus exists a need for a method to generate diverse populations of heteromeric receptors which mimics 35 the natural immune system, which is fast and efficient and

results in only desired combinations without loss of diversity. The present invention satisfies these needs and provides related advantages as well.

SUMMARY OF THE INVENTION

5 The invention relates to a plurality of cells containing diverse combinations of first and second DNA sequences encoding first and second polypeptides which form a heteromeric receptor, said heteromeric receptors being expressed on the surface of a cell, preferably one which
10 produces filamentous bacteriophage, such as M13. Vectors, cloning systems and methods of making and screening the heteromeric receptors are also provided.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic diagram of the two vectors used for surface expression library construction from heavy and light chain libraries. M13IX30 (Figure 1A) is the vector used to clone the heavy chain sequences (open box). The single-headed arrow represents the Lac p/o expression sequences and the double-headed arrow represents the portion of M13IX30 which is to be combined with M13IX11. The amber stop codon and relevant restriction sites are also shown. M13IX11 (Figure 1B) is the vector used to clone the light chain sequences (hatched box). Thick lines represent the pseudo-wild type (gVIII) and wild type (gVIII) gene VIII sequences. The double-headed arrow represents the portion of M13IX11 which is to be combined with M13IX30. Relevant restriction sites are also shown. Figure 1C shows the joining of vector population from heavy and light chain libraries to form the functional surface expression vector M13IXHL. Figure 1D shows the generation of a surface expression library in a non-suppressor strain and the production of phage. The phage are used to infect a suppressor strain (Figure 1E) for surface expression and

screening of the library.

Figure 2 is the nucleotide sequence of M13IX30 (SEQ ID NO: 1).

5 Figure 3 is the nucleotide sequence of M13IX11 (SEQ ID NO: 2).

Figure 4 is the nucleotide sequence of M13IX34 (SEQ ID NO: 3).

Figure 5 is the nucleotide sequence of M13IX13 (SEQ ID NO: 4).

10 Figure 6 is the nucleotide sequence of M13IX60 (SEQ ID NO: 5).

DETAILED DESCRIPTION OF THE INVENTION

This invention is directed to simple and efficient methods to generate a large repertoire of diverse 15 combinations of heteromeric receptors. The method is advantageous in that only proper combinations of vector portions are randomly brought together for the coexpression of different DNA sequences without loss of population size or diversity. The receptors can be expressed on the 20 surface of cells, such as those producing filamentous bacteriophage, which can be screened in large numbers. The nucleic acid sequences encoding the receptors be readily characterized because the filamentous bacteriophage produce single strand DNA for efficient sequencing and mutagenesis 25 methods. The heteromeric receptors so produced are useful in an unlimited number of diagnostic and therapeutic procedures.

In one embodiment, two populations of diverse heavy (Hc) and light (Lc) chain sequences are synthesized by

polymerase chain reaction (PCR). These populations are cloned into separate M13-based vector containing elements necessary for expression. The heavy chain vector contains a gene VIII (gVIII) coat protein sequence so that 5 translation of the Hc sequences produces gVIII-Hc fusion proteins. The populations of two vectors are randomly combined such that only the vector portions containing the Hc and Lc sequences are joined into a single circular vector. The combined vector directs the coexpression of 10 both Hc and Lc sequences for assembly of the two polypeptides and surface expression on M13. A mechanism also exists to control the expression of gVIII-Hc fusion proteins during library construction and screening.

As used herein, the term "heteromeric receptors" 15 refers to proteins composed of two or more subunits which together exhibit binding activity toward particular molecule. It is understood that the term includes the subunit fragments so long as assembly of the polypeptides and function of the assembled complex is retained. 20 Heteromeric subunits include, for example, antibodies and fragments thereof such as Fab and (Fab)₂ portions, T cell receptors, integrins, hormone receptors and transmitter receptors.

As used herein, the term "preselected molecule" refers 25 to a molecule which is chosen from a number of choices. The molecule can be, for example, a protein or peptide, or an organic molecule such as a drug. Benzodiazepam is a specific example of a preselected molecule.

As used herein, the term "coexpression" refers to the 30 expression of two or more nucleic acid sequences usually expressed as separate polypeptides. For heteromeric receptors, the coexpressed polypeptides assemble to form the heteromer. Therefore, "expression elements" as used herein, refers to sequences necessary for the

transcription, translation, regulation and sorting of the expressed polypeptides which make up the heteromeric receptors. The term also includes the expression of two subunit polypeptides which are linked but are able to 5 assemble into a heteromeric receptor. A specific example of coexpression of linked polypeptides is where Hc and Lc polypeptides are expressed with a flexible peptide or polypeptide linker joining the two subunits into a single chain. The linker is flexible enough to allow association 10 of Hc and Lc portions into a functional Fab fragment.

The invention provides for a composition of matter comprising a plurality of prokaryotic cells containing diverse combinations of first and second DNA sequences encoding first and second polypeptides which form a 15 heteromeric receptor exhibiting binding activity toward a preselected molecule, said heteromeric receptors being expressed on the surface of filamentous bacteriophage.

DNA sequences encoding the polypeptides of heteromeric receptors are obtained by methods known to one 20 skilled in the art. Such methods include, for example, cDNA synthesis and polymerase chain reaction (PCR). The need will determine which method or combinations of methods is to be used to obtain the desired populations of sequences. Expression can be performed in any compatible 25 vector/host system. Such systems include, for example, plasmids or phagemids in prokaryotes such as E. coli, yeast systems and other eucaryotic systems such as mammalian cells, but will be described herein in context with its presently preferred embodiment, i.e. expression on the 30 surface of filamentous bacteriophage. Filamentous bacteriophage include, for example, M13, f1 and fd. Additionally, the heteromeric receptors can also be expressed in soluble or secreted form depending on the need and the vector/host system employed.

Expression of heteromeric receptors such as antibodies or functional fragments thereof on the surface of M13 can be accomplished, for example, using the vector system shown in Figure 1. Construction of the vectors enabling one of ordinary skill to make them are explicitly set out in Example I. The complete nucleotide sequences are given in Figures 2 and 3 (SEQ ID NOS: 1 and 2). This system produces randomly combined populations of heavy (Hc) and light (Lc) chain antibody fragments functionally linked to expression elements. The Hc polypeptide is produced as a fusion protein with the M13 coat protein encoded by gene VIII. The gVIII-Hc fusion protein therefore anchors the assembled Hc and Lc polypeptides on the surface of M13. The diversity of Hc and Lc combinations obtained by this system can be 5×10^7 or greater. Diversity of less than 5×10^7 can also be obtained and will be determined by the need and type of heteromeric receptor to be expressed.

Populations of Hc and Lc encoding sequences to be combined into a vector for coexpression are each cloned into separate vectors. For the vectors shown in Figure 1, diverse populations of sequences encoding Hc polypeptides are cloned into M13IX30 (SEQ ID NO: 1). Sequences encoding Lc polypeptides are cloned into M13IX11 (SEQ ID NO: 2). The populations are inserted between the Xho I-Spe I or Stu I restriction enzyme sites in M13IX30 and between the Sac I-Xba I or Eco RV sites in M13IX11 (Figures 1A and B, respectively).

The populations of Hc and Lc sequences inserted into the vectors can be synthesized with appropriate restriction recognition sequences flanking opposite ends of the encoding sequences but this is not necessary. The sites allow annealing and ligation in-frame with expression elements of these sequences into a double-stranded vector restricted with the appropriate restriction enzyme. Alternatively, and a preferred embodiment, the Hc and Lc

sequences can be inserted into the vector without restriction of the DNA. This method of cloning is beneficial because naturally encoded restriction enzyme sites may be present within the sequences, thus, causing 5 destruction of the sequence when treated with a restriction enzyme. For cloning without restriction, the sequences are treated briefly with a 3' to 5' exonuclease such as T4 DNA polymerase or exonuclease III. A 5' to 3' exonuclease will also accomplish the same function. The protruding 5' 10 termini which remains should be complementary to single stranded overhangs within the vector which remain after restriction at the cloning site and treatment with exonuclease. The exonuclease treated inserts are annealed with the restricted vector by methods known to one skilled 15 in the art. The exonuclease method decreases background and is easier to perform.

The vector used for Hc populations, M13IX30 (Figure 1A; SEQ ID NO: 1) contains, in addition to expression elements, a sequence encoding the pseudo-wild type gVIII 20 product downstream and in frame with the cloning sites. This gene encodes the wild type M13 gVIII amino acid sequence but has been changed at the nucleotide level to reduce homologous recombination with the wild type gVIII contained on the same vector. The wild type gVIII is 25 present to ensure that at least some functional, non-fusion coat protein will be produced. The inclusion of a wild type gVIII therefore reduces the possibility of non-viable phage production and biological selection against certain peptide fusion proteins. Differential regulation of the 30 two genes can also be used to control the relative ratio of the pseudo and wild type proteins.

Also contained downstream and in frame with the cloning sites is an amber stop codon. The stop codon is located between the inserted Hc sequences and the gVIII 35 sequence and is in frame. As was the function of the wild

type gVIII, the amber stop codon also reduces biological selection when combining vector portions to produce functional surface expression vectors. This is accomplished by using a non-suppressor (sup 0) host strain 5 because the non-suppressor strains will terminate expression after the Hc sequences but before the pseudo gVIII sequences. Therefore, the pseudo gVIII will essentially never be expressed on the phage surface under these circumstances. Instead, only soluble Hc polypeptides 10 will be produced. Expression in a non-suppressor host strain can be advantageously utilized when one wishes to produce large populations of antibody fragments. Stop codons other than amber, such as opal and ochre, or molecular switches, such as inducible repressor elements, 15 can also be used to unlink peptide expression from surface expression.

The vector used for Lc populations, M13IX11 (SEQ ID NO: 2), contains necessary expression elements and cloning sites for the Lc sequences, Figure 1B. As with M13IX30, 20 upstream and in frame with the cloning sites is a leader sequence for sorting to the phage surface. Additionally, a ribosome binding site and Lac Z promoter/operator elements are also present for transcription and translation of the DNA sequences.

25 Both vectors contain two pairs of Mlu I-Hind III restriction enzyme sites (Figures 1A and B) for joining together the Hc and Lc encoding sequences and their associated vector sequences. Mlu I and Hind III are non-compatible restriction sites. The two pairs are 30 symmetrically orientated about the cloning site so that only the vector portions containing the sequences to be expressed are exactly combined into a single vector. The two pairs of sites are oriented identically with respect to one another on both vectors and the DNA between the two 35 sites must be homologous enough between both vectors to

allow annealing. This orientation allows cleavage of each circular vector into two portions and combination of essential components within each vector into a single circular vector where the encoded polypeptides can be 5 coexpressed (Figure 1C).

Any two pairs of restriction enzyme sites can be used so long as they are symmetrically orientated about the cloning site and identically orientated on both vectors. The sites within each pair, however, should be non-10 identical or able to be made differentially recognized as a cleavage substrate. For example, the two pairs of restriction sites contained within the vectors shown in Figure 1 are Mlu I and Hind III. The sites are differentially cleavable by Mlu I and Hind III 15 respectively. One skilled in the art knows how to substitute alternative pairs of restriction enzyme sites for the Mlu I-Hind III pairs described above. Also, instead of two Hind III and two Mlu I sites, a Hind III and Not I site can be paired with a Mlu I and a Sal I site, for 20 example.

The combining step randomly brings together different Hc and Lc encoding sequences within the two diverse populations into a single vector (Figure 1C; M13IXHL). The vector sequences donated from each independent vector, 25 M13IX30 and M13IX11, are necessary for production of viable phage. Also, since the pseudo gVIII sequences are contained in M13IX30, coexpression of functional antibody fragments as Lc associated gVIII-Hc fusion proteins cannot be accomplished on the phage surface until the vector 30 sequences are linked as shown in M13IXHL.

The combining step is performed by restricting each population of Hc and Lc containing vectors with Mlu I and Hind III, respectively. The 3' termini of each restricted vector population is digested with a 3' to 5' exonuclease

as described above for inserting sequences into the cloning sites. The vector populations are mixed, allowed to anneal and introduced into an appropriate host. A non-suppressor host (Figure 1D) is preferably used during initial 5 construction of the library to ensure that sequences are not selected against due to expression as fusion proteins. Phage isolated from the library constructed in a non-suppressor strain can be used to infect a suppressor strain for surface expression of antibody fragments.

10 A method for selecting a heteromeric receptor exhibiting binding activity toward a preselected molecule from a population of diverse heteromeric receptors, comprising: (a) operationally linking to a first vector a first population of diverse DNA sequences encoding a 15 diverse population of first polypeptides, said first vector having two pairs of restriction sites symmetrically oriented about a cloning site; (b) operationally linking to a second vector a second population of diverse DNA sequences encoding a diverse population of second 20 polypeptides, said second vector having two pairs of restriction sites symmetrically oriented about a cloning site in an identical orientation to that of the first vector; (c) combining the vector products of step (a) and (b) under conditions which allow only the operational 25 combination of vector sequences containing said first and second DNA sequences; (d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of first and second DNA sequences; and (e) determining the heteromeric 30 receptors which bind to said preselected molecule. The invention also provides for determining the nucleic acid sequences encoding such polypeptides as well.

35 Surface expression of the antibody library is performed in an amber suppressor strain. As described above, the amber stop codon between the Hc sequence and the

gVIII sequence unlinks the two components in a non-suppressor strain. Isolating the phage produced from the non-suppressor strain and infecting a suppressor strain will link the Hc sequences to the gVIII sequence during expression (Figure 1E). Culturing the suppressor strain after infection allows the coexpression on the surface of M13 of all antibody species within the library as gVIII fusion proteins (gVIII-Fab fusion proteins). Alternatively, the DNA can be isolated from the non-suppressor strain and then introduced into a suppressor strain to accomplish the same effect.

The level of expression of gVIII-Fab fusion proteins can additionally be controlled at the transcriptional level. Both polypeptides of the gVIII-Fab fusion proteins are under the inducible control of the Lac Z promoter/operator system. Other inducible promoters can work as well and are known by one skilled in the art. For high levels of surface expression, the suppressor library is cultured in an inducer of the Lac Z promoter such as isopropylthio- β -galactoside (IPTG). Inducible control is beneficial because biological selection against non-functional gVIII-Fab fusion proteins can be minimized by culturing the library under non-expressing conditions. Expression can then be induced only at the time of screening to ensure that the entire population of antibodies within the library are accurately represented on the phage surface. Also, this can be used to control the valency of the antibody on the phage surface.

The surface expression library is screened for specific Fab fragments which bind preselected molecules by standard affinity isolation procedures. Such methods include, for example, panning, affinity chromatography and solid phase blotting procedures. Panning as described by Parmley and Smith, Gene 73:305-318 (1988), which is incorporated herein by reference, is preferred because high

titers of phage can be screened easily, quickly and in small volumes. Furthermore, this procedure can select minor Fab fragments species within the population, which otherwise would have been undetectable, and amplified to substantially homogenous populations. The selected Fab fragments can be characterized by sequencing the nucleic acids encoding the polypeptides after amplification of the phage population.

The following examples are intended to illustrate but not limit the invention.

EXAMPLE I

Construction, Expression and Screening of Antibody Fragments on the Surface of M13

This example shows the synthesis of a diverse population of heavy (Hc) and light (Lc) chain antibody fragments and their expression on the surface of M13 as gene VIII-Fab fusion proteins. The expressed antibodies derive from the random mixing and coexpression of a Hc and Lc pair. Also demonstrated is the isolation and characterization of the expressed Fab fragments which bind benzodiazepam (BDP) and their corresponding nucleotide sequence.

Isolation of mRNA and PCR Amplification of Antibody Fragments

The surface expression library is constructed from mRNA isolated from a mouse that had been immunized with KLH-coupled benzodiazepam (BDP). BDP was coupled to keyhole limpet hemocyanin (KLH) using the techniques described in Antibodies: A Laboratory Manual, Harlow and Lane, eds., Cold Spring Harbor, New York (1988), which is incorporated herein by reference. Briefly, 10.0 milligrams (mg) of keyhole limpet hemocyanin and 0.5 mg of BDP with a

glutaryl spacer arm N-hydroxysuccinimide linker appendages. Coupling was performed as in Jonda et al., Science, 241:1188 (1988), which is incorporated herein by reference. The KLH-BDP conjugate was removed by gel filtration 5 chromatography through Sephadex G-25.

The KLH-BDP conjugate was prepared for injection into mice by adding 100 μ g of the conjugate to 250 μ l of phosphate buffered saline (PBS). An equal volume of complete Freund's adjuvant was added and emulsified the 10 entire solution for 5 minutes. Mice were injected with 300 μ l of the emulsion. Injections were given subcutaneously at several sites using a 21 gauge needle. A second immunization with BDP was given two weeks later. This 15 injection was prepared as follows: 50 μ g of BDP was diluted in 250 μ l of PBS and an equal volume of alum was mixed with the solution. The mice were injected intraperitoneally with 500 μ l of the solution using a 23 gauge needle. One month later the mice were given a final 20 injection of 50 μ g of the conjugate diluted to 200 μ l in PBS. This injection was given intravenously in the lateral tail vein using a 30 gauge needle. Five days after this final injection the mice were sacrificed and total cellular RNA was isolated from their spleens.

Total RNA was isolated from the spleen of a single 25 mouse immunized as described above by the method of Chomczynski and Sacchi, Anal. Biochem., 162:156-159 (1987), which is incorporated herein by reference. Briefly, immediately after removing the spleen from the immunized mouse, the tissue was homogenized in 10 ml of a denaturing 30 solution containing 4.0 M guanine isothiocyanate, 0.25 M sodium citrate at pH 7.0, and 0.1 M 2-mercaptoethanol using a glass homogenizer. One ml of sodium acetate at a concentration of 2 M at pH 4.0 was mixed with the homogenized spleen. One ml of saturated phenol was also 35 mixed with the denaturing solution containing the

homogenized spleen. Two ml of a chloroform:isoamyl alcohol (24:1 v/v) mixture was added to this homogenate. The homogenate was mixed vigorously for ten seconds and maintained on ice for 15 minutes. The homogenate was then 5 transferred to a thick-walled 50 ml polypropylene centrifuge tube (Fisher Scientific Company, Pittsburgh, PA). The solution was centrifuged at 10,000 x g for 20 minutes at 4°C. The upper RNA-containing aqueous layer was transferred to a fresh 50 ml polypropylene centrifuge tube 10 and mixed with an equal volume of isopropyl alcohol. This solution was maintained at -20°C for at least one hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for twenty minutes at 4°C. The pelleted total cellular RNA was 15 collected and dissolved in 3 ml of the denaturing solution described above. Three mls of isopropyl alcohol was added to the resuspended total cellular RNA and vigorously mixed. This solution was maintained at -20°C for at least 1 hour to precipitate the RNA. The solution containing the 20 precipitated RNA was centrifuged at 10,000 x g for ten minutes at 4°C. The pelleted RNA was washed once with a solution containing 75% ethanol. The pelleted RNA was dried under vacuum for 15 minutes and then resuspended in dimethyl pyrocarbonate (DEPC) treated (DEPC-H₂O) H₂O.

25 Poly A' RNA for use in first strand cDNA synthesis was prepared from the above isolated total RNA using a spin-column kit (Pharmacia, Piscataway, NJ) as recommended by the manufacturer. The basic methodology has been described by Aviv and Leder, Proc. Natl. Acad. Sci., USA, 69:1408-30 1412 (1972), which is incorporated herein by reference. Briefly, one half of the total RNA isolated from a single immunized mouse spleen prepared as described above was resuspended in one ml of DEPC-treated dH₂O and maintained at 65°C for five minutes. One ml of 2x high salt loading 35 buffer (100 mM Tris-HCL at pH 7.5, 1 M sodium chloride, 2.0 mM disodium ethylene diamine tetraacetic acid (EDTA) at pH

8.0, and 0.2% sodium dodecyl sulfate (SDS)) was added to the resuspended RNA and the mixture was allowed to cool to room temperature. The mixture was then applied to an oligo-dT (Collaborative Research Type 2 or Type 3 Bedford, MA) column that was previously prepared by washing the oligo-dT with a solution containing 0.1 M sodium hydroxide and 5 mM EDTA and then equilibrating the column with DEPC-treated dH₂O. The eluate was collected in a sterile polypropylene tube and reapplied to the same column after 5 heating the eluate for 5 minutes at 65°C. The oligo dT column was then washed with 2 ml of high salt loading buffer consisting of 50 mM Tris-HCL at pH 7.5, 500 mM sodium chloride, 1 mM EDTA at pH 8.0 and 0.1% SDS. The oligo dT column was then washed with 2 ml of 1 X medium 10 salt buffer (50 mM Tris-HCL at pH 7.5, 100 mM sodium chloride, 1 mM EDTA at pH 8.0 and 0.1% SDS). The mRNA was eluted with 1 ml of buffer consisting of 10 mM Tris-HCL at pH 7.5, 1 mM EDTA at pH 8.0 and 0.05% SDS. The messenger RNA was purified by extracting this solution with 15 phenol/chloroform followed by a single extraction with 100% chloroform, ethanol precipitated and resuspended in DEPC treated dH₂O.

In preparation for PCR amplification, mRNA was used as a template for cDNA synthesis. In a typical 250 µl reverse 25 transcription reaction mixture, 5-10 µg of spleen mRNA in water was first annealed with 500 ng (0.5 pmol) of either the 3' V_h primer (primer 12, Table I) or the 3' V_l primer (primer 9, Table II) at 65°C for 5 minutes. Subsequently, the mixture was adjusted to contain 0.8 mM dATP, 0.8 mM 30 dCTP, 0.8 mM dGTP, 0.8 mM dTTP, 100 mM Tris-HCL (pH 8.6), 10 mM MgCl₂, 40 mM KCl, and 20 mM 2-ME. Moloney-Murine Leukemia Virus (Bethesda Research Laboratories (BRL), Gaithersburg, MD) Reverse transcriptase, 26 units, was added and the solution was incubated for 1 hour at 40°C. 35 The resultant first strand cDNA was phenol extracted, ethanol precipitated and then used in the polymerase chain

reaction (PCR) procedures described below for amplification of heavy and light chain sequences.

Primers used for amplification of heavy chain Fd fragments for construction of the M13IX30 library is shown 5 in Table I. Amplification was performed in eight separate reactions, as described by Saiki et al., Science, 239:487-491 (1988), which is incorporated herein by reference, each reaction containing one of the 5' primers (primers 2 to 9; SEQ ID NOS: 7 through 14, respectively) and one of the 3' 10 primers (primer 12; SEQ ID NO: 17) listed in Table I. The remaining 5' primers, used for amplification in a single reaction, are either a degenerate primer (primer 1; SEQ ID NO: 6) or a primer that incorporates inosine at four degenerate positions (primer 10; SEQ ID NO: 15). The 15 remaining 3' primer (primer 11; SEQ ID NO: 16) was used to construct Fv fragments. The underlined portion of the 5' primers incorporates an Xho I site and that of the 3' primer an Spe I restriction site for cloning the amplified fragments into the M13IX30 vector in a predetermined 20 reading frame for expression.

TABLE I
HEAVY CHAIN PRIMERS

		CC G G	T
25	1)	5' - AGGT A CT <u>CTCGAGTC</u> GG - 3'	
		GA A T	A
	2)	5' - AGGTCCAGCTG <u>CTCGAGT</u> CTGG - 3'	
	3)	5' - AGGTCCAGCT <u>CTCGAGTC</u> AGG - 3'	
	4)	5' - AGGTCCAGCT <u>CTCGAGT</u> CTGG - 3'	
	5)	5' - AGGTCCAGCT <u>CTCGAGTC</u> AGG - 3'	
30	6)	5' - AGGTCCA <u>ACTGCTCGAGT</u> CTGG - 3'	
	7)	5' - AGGTCCA <u>ACTGCTCGAGTC</u> AGG - 3'	
	8)	5' - AGGTCCA <u>ACTTCTCGAGT</u> CTGG - 3'	

9) 5' - AGGTCCAACTTCTCGAGTCAGG - 3'

10) 5' - AGGTIIIAICTICTCGAGTC ^T
A GG - 3'

5 11) 5' - CTATTAACTAGTAACGGTAACAGT -
GGTGCCTGCCCA - 3'

12) 5' - AGGCTTACTAGTACAATCCCTGG -
GCACAAAT - 3'

Primers used for amplification of mouse kappa light
 10 chain sequences for construction of the M13IX11 library are
 shown in Table II. These primers were chosen to contain
 restriction sites which were compatible with vector and not
 present in the conserved sequences of the mouse light chain
 mRNA. Amplification was performed as described above in
 15 five separate reactions, each containing one of the 5'
 primers (primers 3 to 7; SEQ ID NOS: 20 through 24,
 respectively) and one of the 3' primers (primer 9; SEQ ID
 NO: 26) listed in Table II. The remaining 3' primer
 (primer 8; SEQ ID NO: 25) was used to construct Fv
 20 fragments. The underlined portion of the 5' primers
 depicts a Sac I restriction site and that of the 3' primers
 an Xba I restriction site for cloning of the amplified
 fragments into the M13IX11 vector in a predetermined
 reading frame for expression.

25

TABLE II
LIGHT CHAIN PRIMERS

1) 5' - CCAGTTCCGAGCTCGTTGTGACTCAGGAATCT - 3'
 2) 5' - CCAGTTCCGAGCTCGTGTGACTCAGCCGCC - 3'
 3) 5' - CCAGTTCCGAGCTCGTGTGACTCAGCTCCA - 3'
 30 4) 5' - CCAGTTCCGAGCTCCAGATGACCCAGTCTCCA - 3'
 5) 5' - CCAGATGTGAGCTCGTGATGACCCAGACTCCA - 3'
 6) 5' - CCAGATGTGAGCTCGTCATGACCCAGTCTCCA - 3'
 7) 5' - CCAGTTCCGAGCTCGTGATGACACAGTCTCCA - 3'
 8) 5' - GCAGCATTCTAGAGTTCAGCTCCAGCTTGCC - 3'
 35 9) 5' - GCGCCGTCTAGAATTAAACACTCATTCCCTGTTGAA - 3'

PCR amplification for heavy and light chain fragments was performed in a 100 μ l reaction mixture containing the above described products of the reverse transcription reaction (\approx 5 μ g of the cDNA-RNA hybrid), 300 nmol of 3' V_h primer (primer 12, Table I; SEQ ID NO: 17), and one of the 5' V_h primers (primers 2-9, Table I; SEQ ID NOS: 7 through 14, respectively) for heavy chain amplification, or, 300 nmol of 3' V_l primer (primer 9, Table II; SEQ ID NO: 26), and one of the 5' V_l primers (primers 3-7, Table II; SEQ ID NOS: 20 through 24, respectively) for each light chain amplification, a mixture of dNTPs at 200 mM, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 15 mM MgCl₂, 0.1% gelatin, and 2 units of *Thermus aquaticus* DNA polymerase. The reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle involved denaturation at 92°C for 1 minute, annealing at 52°C for 2 minutes, and elongation at 72°C for 1.5 minutes. The amplified samples were extracted twice with phenol/CHCl₃ and once with CHCl₃, ethanol-precipitated, and stored at -70°C in 10 mM Tris-HCl, pH 7.5 1 mM EDTA. The resultant products were used in constructing the M13IX30 and M13IX11 libraries (see below).

Vector Construction

Two M13-based vectors, M13IX30 (SEQ ID NO: 1) and 25 M13IX11 (SEQ ID NO: 2), were constructed for the cloning and propagation of Hc and Lc populations of antibody fragments, respectively. The vectors were constructed to facilitate the random joining and subsequent surface expression of antibody fragment populations.

30 M13IX30 (SEQ ID NO: 1), or the Hc vector, was constructed to harbor diverse populations of Hc antibody fragments. M13mp19 (Pharmacia, Piscataway, NJ) was the starting vector. This vector was modified to contain, in addition to the encoded wild type M13 gene VIII: (1) a

pseudo-wild type gene VIII sequence with an amber stop codon between it and the restriction sites for cloning oligonucleotides; (2) Stu I restriction site for insertion of sequences by hybridization and, Spe I and Xho I restriction sites in-frame with the pseudo-wild type gene VIII for cloning Hc sequences; (3) sequences necessary for expression, such as a promoter, signal sequence and translation initiation signals; (4) two pairs of Hind III-Mlu I sites for random joining of Hc and Lc vector portions, and (5) various other mutations to remove redundant restriction sites and the amino terminal portion of Lac Z.

Construction of M13IX30 was performed in four steps. In the first step, an M13-based vector containing the 15 pseudo gVIII and various other mutations was constructed, M13IX01F. The second step involved the construction of a small cloning site in a separate M13mp18 vector to yield M13IX03. This vector was then expanded to contain expression sequences and restriction sites for Hc sequences 20 to form M13IX04B. The fourth and final step involved the incorporation of the newly constructed sequences in M13IX04B into M13IX01F to yield M13IX30.

Construction of M13IX01F first involved the generation of a pseudo wild-type gVIII sequence for surface expression 25 of antibody fragments. The pseudo-wild type gene encodes the identical amino acid sequence as that of the wild type gene; however, the nucleotide sequence has been altered so that only 63% identity exists between this gene and the encoded wild type gene VIII. Modification of the gene VIII 30 nucleotide sequence used for surface expression reduces the possibility of homologous recombination with the wild type gene VIII contained on the same vector. Additionally, the wild type M13 gene VIII was retained in the vector system to ensure that at least some functional, non-fusion coat 35 protein would be produced. The inclusion of wild type gene

VIII facilitates the growth of phage under conditions where there is surface expression of the polypeptides and therefore reduces the possibility of non-viable phage production from the fusion genes.

5 The pseudo-wild type gene VIII was constructed by chemically synthesizing a series of oligonucleotides which encode both strands of the gene. The oligonucleotides are presented in Table III.

TABLE IIIPseudo-Wild Type Gene VIII Oligonucleotide Series

	<u>Top Strand Oligonucleotides</u>	<u>Sequence (5' to 3')</u>
5	VIII 03	GATCC TAG GCT GAA GGC GAT GAC CCT GCT AAG GCT GC
10	VIII 04	A TTC AAT AGT TTA CAG GCA AGT GCT ACT GAG TAC A
15	VIII 05	TT GGC TAC GCT TGG GCT ATG GTA GTA GTT ATA GTT GGT GCT ACC ATA GGG ATT AAA TTA TTC AAA AAG TT T ACG AGC AAG GCT TCT TA
	<u>Bottom Strand Oligonucleotides</u>	
20	VIII 08	AGC TTA AGA AGC CTT GCT CGT AAA CTT TTT GAA TAA TTT
25	VIII 09	AAT CCC TAT GGT AGC ACC AAC TAT AAC TAC TAC CAT AGC CCA AGC GTA GCC AAT GTA CTC AGT AGC ACT TG C CTG TAA ACT ATT GAA TGC AGC CTT AGC AGG GTC ATC GCC TTC AGC CTA G
30	VIII 10	
	VIII 11	
	VIII 12	
		Except for the terminal oligonucleotides VIII 03 (SEQ ID NO: 27) and VIII 08 (SEQ ID NO: 32), the above oligonucleotides (oligonucleotides VIII 04-07 (SEQ ID NOS: 28 through 31, respectively) and VIII 09-12 (SEQ ID NOS: 33

through 36, respectively)) were mixed at 200 ng each in 10 μ l final volume, phosphorylated with T4 polynucleotide Kinase (Pharmacia) and 1 mM ATP at 37°C for 1 hour, heated to 70°C for 5 minutes, and annealed into double-stranded 5 form by heating to 65°C for 3 minutes, followed by cooling to room temperature over a period of 30 minutes. The reactions were treated with 1.0 U of T4 DNA ligase (BRL) and 1 mM ATP at room temperature for 1 hour, followed by heating to 70°C for 5 minutes. Terminal oligonucleotides 10 were then annealed to the ligated oligonucleotides. The annealed and ligated oligonucleotides yielded a double-stranded DNA flanked by a Bam HI site at its 5' end and by a Hind III site at its 3' end. A translational stop codon (amber) immediately follows the Bam HI site. The gene VIII 15 sequence begins with the codon GAA (Glu) two codons 3' to the stop codon. The double-stranded insert was cloned in frame with the Eco RI and Sac I sites within the M13 polylinker. To do so, M13mp19 was digested with Bam HI (New England Biolabs, Beverley, MA) and Hind III (New 20 England Biolabs) and combined at a molar ratio of 1:10 with the double-stranded insert. The ligations were performed at room temperature overnight in 1X ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM MgCl₂, 20 mM DTT, 1 mM ATP, 50 μ g/ml BSA) containing 1.0 U of T4 DNA ligase (New England 25 Biolabs). The ligation mixture was transformed into a host and screened for positive clones using standard procedures in the art.

Several mutations were generated within the construct to yield functional M13I..01F. The mutations were generated 30 using the method of Kunkel et al., Meth. Enzymol. 154:367-382 (1987), which is incorporated herein by reference, for site-directed mutagenesis. The reagents, strains and protocols were obtained from a Bio Rad Mutagenesis kit (Bio Rad, Richmond, CA) and mutagenesis was performed as 35 recommended by the manufacturer.

Two Fok I sites were removed from the vector as well as the Hind III site at the end of the pseudo gene VIII sequence using the mutant oligonucleotides 5'-CATTTCAGATGGCTAGA-3' (SEQ ID NO: 37) and 5'-
5 TAGCATTAACGTCCAATA-3' (SEQ ID NO: 38). New Hind III and Mlu I sites were also introduced at position 3919 and 3951 of M13IX01F. The oligonucleotides used for this mutagenesis had the sequences 5'-ATATATTTAGTAAGCTTCATCTTCT-3' (SEQ ID NO: 39) and 5'-
10 GACAAAGAACGGTGAAACTT-3' (SEQ ID NO: 40), respectively. The amino terminal portion of Lac Z was deleted by oligonucleotide-directed mutagenesis using the mutant oligonucleotide 5'-GCGGGCCTTCGCTATTGCTTAAGAAGCCTTGCT-3' (SEQ ID NO: 41). In constructing the above mutations, all 15 changes made in a M13 coding region were performed such that the amino acid sequence remained unaltered. The resultant vector, M13IX01F, was used in the final step to construct M13IX30 (see below).

In the second step, M13mp18 was mutated to remove the 20 5' end of Lac Z up to the Lac i binding site and including the Lac Z ribosome binding site and start codon. Additionally, the polylinker was removed and a Mlu I site was introduced in the coding region of Lac Z. A single oligonucleotide was used for these mutagenesis and had the 25 sequence 5'-AAACGACGGCCAGTGCCAAAGTGACCGCGTGTGAAATTGTTATCC-3' (SEQ ID NO: 42). Restriction enzyme sites for Hind III and Eco RI were introduced downstream of the Mlu I site using the oligonucleotide 5'-GGCGAAAGGAAATTCTGCAAGGCGATTAAGCTTGGG TAACGCC-3' (SEQ ID NO. 43). These modifications of M13mp18 30 yielded the precursor vector M13IX03.

The expression sequences and cloning sites were introduced into M13IX03 by chemically synthesizing a series of oligonucleotides which encode both strands of the desired sequence. The oligonucleotides are presented in 35 Table IV.

TABLE IV
M13IX30 Oligonucleotide Series

Top Strand <u>Oligonucleotides</u>		<u>Sequence (5' to 3')</u>
5	084	GGCGTTACCCAAGCTTGACATGGAGAAAATAAAG
	027	TGAAACAAAGCACTATTGCACTGGCACTCTTACCGT TACCGT
	028	TACTGTTACCCCTGTGACAAAAGCCGCCAGGTCC AGCTGC
10	029	TCGAGTCAGGCCTATTGTGCCAGGGATTGTACTAG TGGATCCG
Bottom <u>Oligonucleotides</u>		<u>Sequence (5' to 3')</u>
	085	TGGCGAAAGGAATTGGATCCACTAGTACAATCCCTG
15	031	GGCACAAATAGGCCTGACTCGAGCAGCTGGACCAGGGCG GCTT
	032	TTGTCACAGGGTAAACAGTAACGGTAACGGTAAGTGT GCCA
20	033	GTGCAATAGTGCTTGTTCACTTATTTCTCCATGT ACAA

The above oligonucleotides of Table IV, except for the terminal oligonucleotides 084 (SEQ ID NO: 44) and 085 (SEQ ID NO: 48), were mixed, phosphorylated, annealed and ligated to form a double-stranded insert as described in Example I. However, instead of cloning directly into the intermediate vector the insert was first amplified by PCR. The terminal oligonucleotides were used as primers for PCR. Oligonucleotide 084 (SEQ ID NO: 44) contains a Hind III site, 10 nucleotides internal to its 5' end and oligonucleotide 085 (SEQ ID NO: 48) has an Eco RI site at its 5' end. Following amplification, the products were restricted with Hind III and Eco RI and ligated, as described in Example I, into the polylinker of M13mp18 digested with the same two enzymes. The resultant double

stranded insert contained a ribosome binding site, a translation initiation codon followed by a leader sequence and three restriction enzyme sites for cloning random oligonucleotides (Xho I, Stu I, Spe I). The intermediate 5 vector was named M13IX04.

During cloning of the double-stranded insert, it was found that one of the GCC codons in oligonucleotides 028 and its complement in 031 was deleted. Since this deletion did not affect function, the final construct is missing one 10 of the two GCC codons. Additionally, oligonucleotide 032 (SEQ ID NO: 50) contained a GTG codon where a GAG codon was needed. Mutagenesis was performed using the oligonucleotide 5'-TAACGGTAAGAGTGCCAGTGC-3' (SEQ ID NO: 52) to convert the codon to the desired sequence. The 15 resultant vector is named M13IX04B.

The third step in constructing M13IX30 involved inserting the expression and cloning sequences from M13IX04B upstream of the pseudo wild-type gVIII in M13IX01F. This was accomplished by digesting M13IX04B with 20 Dra III and Bam HI and gel isolating the 700 base pair insert containing the sequences of interest. M13IX01F was likewise digested with Dra III and Bam HI. The insert was combined with the double digested vector at a molar ratio of 1:1 and ligated as described in Example I. The sequence 25 of the final construct M13IX30, is shown in Figure 2 (SEQ ID NO: 1). Figure 1A also shows M13IX30 where each of the elements necessary for surface expression of Hc fragments is marked. It should be noted during modification of the vectors, certain sequences differed from the published 30 sequence of M13mp18. The new sequences are incorporated into the sequences recorded herein.

M13IX11 (SEQ ID NO: 2), or the Lc vector, was constructed to harbor diverse populations of Lc antibody fragments. This vector was also constructed from M13mp19

and contains: (1) sequences necessary for expression, such as a promoter, signal sequence and translation initiation signals; (2) Eco RV restriction site for insertion of sequences by hybridization and Sac I and Xba I restriction sites for cloning of Lc sequences; (3) two pairs of Hind III-Mlu I sites for random joining of Hc and Lc vector portions, and (4) various other mutation to remove redundant restriction sites.

The expression, translation initiation signals, cloning sites, and one of the Mlu I sites were constructed by annealing of overlapping oligonucleotides as described above to produce a double-stranded insert containing a 5' Eco RI site and a 3' Hind III site. The overlapping oligonucleotides are shown in Table V and were ligated as a double-stranded insert between the Eco RI and Hind III sites of M13mp18 as described for the expression sequences inserted into M13IX03. The ribosome binding site (AGGAGAC) is located in oligonucleotide 015 and the translation initiation codon (ATG) is the first three nucleotides of oligonucleotide 016 (SEQ ID NO: 55).

TABLE V

Oligonucleotide Series for Construction of
Translation Signals in M13IX11

	<u>Oligonucleotide</u>	<u>Sequence (5' to 3')</u>
5	082	CACC TTCATG AATTC GGC AAG GAGACA GTCAT
	015	AATT C GCC AAG GAG ACA GTC AT
	016	AATG AAA TAC CTA TTG CCT ACG GCA GCC GCT GGA TTG TT
	017	ATTA CTC GCT GCC CAA CCA GCC ATG GCC GAG CTC GTG AT
	018	GACC CAG ACT CCA GATATC CAA CAG GAA TGA GTG TTA AT
	019	TCT AGA ACG CGT C
10	083	TTCAGGTTGAAGC TTA CGC GTT CTA GAA TTA ACA CTC ATT CCTGT
	021	TG GAT ATC TGG AGT CTG GGT CAT CAC GAG CTC GGC CAT G
	022	GC TGG TTG GGC AGC GAG TAA TAA CAA TCC AGC GGC TGC C
	023	GT AGG CAA TAG GTA TTT CAT TAT GAC TGT CCT TGG CG

25 Oligonucleotide 017 (SEQ ID NO: 56) contained a Sac I restriction site 67 nucleotides downstream from the ATG codon. The naturally occurring Eco RI site was removed and new Eco RI and Hind III sites were introduced downstream from the Sac I. Oligonucleotides 5'-TGACTGTCTCCTTGGCGTGTGAAATTGTTA-3' (SEQ ID NO: 63) and 5'-TAACACTCATTCCGGATGGAATTCTGGAGTCTGGGT-3' (SEQ ID NO: 64) were used to generate each of the mutations, respectively. The Lac Z ribosome binding site was removed when the

original Eco RI site in M13mp19 was mutated. Additionally, when the new Eco RI and Hind III sites were generated, a spontaneous 100 bp deletion was found just 3' to these sites. Since the deletion does not affect the function, it 5 was retained in the final vector.

In addition to the above mutations, a variety of other modifications were made to incorporate or remove certain sequences. The Hind III site used to ligate the double-stranded insert was removed with the oligonucleotide 5'-
10 GCCAGTGCCAAAGTGACGCGTTCTA-3' (SEQ ID NO: 65). Second Hind III and Mlu I sites were introduced at positions 3922 and 3952, respectively, using the oligonucleotides 5'-ATATATTTAGTAAGCTTCATCTTCT-3' (SEQ ID NO: 66) for the Hind III mutagenesis and 5'-GACAAAGAACGCGTGAAAACTTT-3' (SEQ ID 15 NO: 67) for the Mlu I mutagenesis. Again, mutations within the coding region did not alter the amino acid sequence.

The sequence of the resultant vector, M13IX11, is shown in Figure 3 (SEQ ID NO: 2). Figure 1B also shows M13IX11 where each of the elements necessary for producing 20 a surface expression library between Lc fragments is marked.

Library Construction

Each population of Hc and Lc sequences synthesized by PCR above are separately cloned into M13IX30 and M13IX11, 25 respectively, to create Hc and Lc libraries.

The Hc and Lc products (5 µg) are mixed, ethanol precipitated and resuspended in 20 µl of NaOAc buffer (33 mM Tris acetate, pH 7.9, 10 mM Mg-acetate, 66 mM K-acetate, 0.5 mM DTT). Five units of T4 DNA polymerase is added and 30 the reactions incubated at 30°C for 5 minutes to remove 3' termini by exonuclease digestion. Reactions are stopped by heating at 70°C for 5 minutes. M13IX30 is digested with

Stu I and M13IX11 is digested with Eco RV. Both vectors are treated with T4 DNA polymerase as described above and combined with the appropriate PCR products at a 1:1 molar ratio at 10 ng/ μ l to anneal in the above buffer at room 5 temperature overnight. DNA from each annealing is electroporated into MK30-3 (Boehringer, Indianapolis, IN), as described below, to generate the Hc and Lc libraries.

E. coli MK30-3 is electroporated as described by Smith et al., Focus 12:38-40 (1990) which is incorporated herein 10 by reference. The cells are prepared by inoculating a fresh colony of MK30-3 into 5 mls of SOB without magnesium (20 g bacto-tryptone, 5 g bacto-yeast extract, 0.584 g NaCl, 0.186 g KCl, dH₂O to 1,000 mls) and grown with vigorous aeration overnight at 37°C. SOB without magnesium 15 (500 ml) is inoculated at 1:1000 with the overnight culture and grown with vigorous aeration at 37°C until the OD₅₅₀ is 0.8 (about 2 to 3 h). The cells are harvested by centrifugation at 5,000 rpm (2,600 x g) in a GS3 rotor (Sorvall, Newtown, CT) at 4°C for 10 minutes, resuspended 20 in 500 ml of ice-cold 10% (v/v) sterile glycerol, centrifuged and resuspended a second time in the same manner. After a third centrifugation, the cells are resuspended in 10% sterile glycerol at a final volume of about 2 ml, such that the OD₅₅₀ of the suspension was 200 to 25 300. Usually, resuspension is achieved in the 10% glycerol that remained in the bottle after pouring off the supernate. Cells are frozen in 40 μ l aliquots in microcentrifuge tubes using a dry ice-ethanol bath and stored frozen at -70°C.

30 Frozen cells are electroporated by thawing slowly on ice before use and mixing with about 10 pg to 500 ng of vector per 40 μ l of cell suspension. A 40 μ l aliquot is placed in an 0.1 cm electroporation chamber (Bio-Rad, Richmond, CA) and pulsed once at 0°C using 4 k Ω parallel 35 resistor 25 μ F, 1.88 KV, which gives a pulse length (τ) of

4 ms. A 10 μ l aliquot of the pulsed cells are diluted into 1 ml SOC (98 mls SOB plus 1 ml of 2 M $MgCl_2$, and 1 ml of 2 M glucose) in a 12- x 75-mm culture tube, and the culture is shaken at 37°C for 1 hour prior to culturing in 5 selective media, (see below).

Each of the libraries are cultured using methods known to one skilled in the art. Such methods can be found in Sanbrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1989, 10 and in Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, New York, 1989, both of which are incorporated herein by reference. Briefly, the above 1 ml library cultures are grown up by diluting 50-fold into 2XYT media (16 g tryptone, 10 g yeast extract, 5 g NaCl) 15 and culturing at 37°C for 5-8 hours. The bacteria are pelleted by centrifugation at 10,000 x g. The supernatant containing phage is transferred to a sterile tube and stored at 4°C.

Double strand vector DNA containing Hc and Lc antibody 20 fragments are isolated from the cell pellet of each library. Briefly, the pellet is washed in TE (10 mM Tris, pH 8.0, 1 mM EDTA) and recollected by centrifugation at 7,000 rpm for 5' in a Sorval centrifuge (Newtown, CT). Pellets are resuspended in 6 mls of 10% Sucrose, 50 mM 25 Tris, pH 8.0. 3.0 ml of 10 mg/ μ l lysozyme is added and incubated on ice for 20 minutes. 12 mls of 0.2 M NaOH, 1% SDS is added followed by 10 minutes on ice. The suspensions are then incubated on ice for 20 minutes after 30 addition of 7.5 mls of 3 M NaOAc, pH 4.6. The samples are centrifuged at 15,000 rpm for 15 minutes at 4°C, RNased and extracted with phenol/chloroform, followed by ethanol precipitation. The pellets are resuspended, weighed and an equal weight of $CsCl_2$ is dissolved into each tube until a density of 1.60 g/ml is achieved. EtBr is added to 600 35 μ g/ml and the double-stranded DNA is isolated by

equilibrium centrifugation in a TV-1665 rotor (Sorval) at 50,000 rpm for 6 hours. These DNAs from each right and left half sublibrary are used to generate forty libraries in which the right and left halves of the randomized 5 oligonucleotides have been randomly joined together.

The surface expression library is formed by the random joining of the Hc containing portion of M13IX30 with the Lc containing portion of M13IX11. The DNAs isolated from each library was digested separately with an excess amount of 10 restriction enzyme. The Lc population (5 μ g) is digested with Hind III. The Hc (5 μ g) population is digested with Mlu I. The reactions are stopped by phenol/chloroform extraction followed by ethanol precipitation. The pellets are washed in 70% ethanol and resuspended in 20 μ l of NaOAc 15 buffer. Five units of T4 DNA polymerase (Pharmacia) is added and the reactions incubated at 30°C for 5 minutes. Reactions are stopped by heating at 70°C for 5 minutes. The Hc and Lc DNAs are mixed to a final concentration of 10 ng each vector/ μ l and allowed to anneal at room temperature 20 overnight. The mixture is electroporated into MK30-3 cells as described above.

Screening of Surface Expression Libraries

Purified phage are prepared from 50 ml liquid cultures of XL1 Blue™ cells (Stratagene, La Jolla, CA) which had 25 been infected at a m.o.i. of 10 from the phage stocks stored at 4°C. The cultures are induced with 2 mM IPTG. Supernatants are cleared by two centrifugations, and the phage are precipitated by adding 1/7.5 volumes of PEG solution (25% PEG-8000, 2.5 M NaCl), followed by incubation 30 at 4°C overnight. The precipitate is recovered by centrifugation for 90 minutes at 10,000 x g. Phage pellets are resuspended in 25 ml of 0.01 M Tris-HCl, pH 7.6, 1.0 mM EDTA, and 0.1% Sarkosyl and then shaken slowly at room temperature for 30 minutes. The solutions are adjusted to

0.5 M NaCl and to a final concentration of 5% polyethylene glycol. After 2 hours at 4°C, the precipitates containing the phage are recovered by centrifugation for 1 hour at 15,000 X g. The precipitates are resuspended in 10 ml of 5 NET buffer (0.1 M NaCl, 1.0 mM EDTA, and 0.01 M Tris-HCl, pH 7.6), mixed well, and the phage repelleted by centrifugation at 170,000 X g for 3 hours. The phage pellets are resuspended overnight in 2 ml of NET buffer and subjected to cesium chloride centrifugation for 18 hours at 10 110,000 X g (3.86 g of cesium chloride in 10 ml of buffer). Phage bands are collected, diluted 7-fold with NET buffer, re-centrifuged at 170,000 X g for 3 hours, resuspended, and stored at 4°C in 0.3 ml of NET buffer containing 0.1 mM sodium azide.

15 The BDP used for panning on streptavidin coated dishes is first biotinylated and then absorbed against UV-inactivated blocking phage (see below). The biotinylating reagents are dissolved in dimethylformamide at a ratio of 2.4 mg solid NHS-SS-Biotin (sulfosuccinimidyl 2- 20 (biotinamido)ethyl-1,3'-dithiopropionate; Pierce, Rockford, IL) to 1 ml solvent and used as recommended by the manufacturer. Small-scale reactions are accomplished by mixing 1 µl dissolved reagent with 43 µl of 1 mg/ml BDP diluted in sterile bicarbonate buffer (0.1 M NaHCO₃, pH 8.6). After 2 hours at 25°C, residual biotinylating reagent is reacted with 500 µl 1 M ethanalamine (pH adjusted to 9 with HCl) for an additional 2 hours. The entire sample is diluted with 1 ml TBS containing 1 mg/ml BSA, concentrated to about 50 µl on a Centricon 30 ultra- 25 filter (Amicon), and washed on the same filter three times with 2 ml TBS and once with 1 ml TBS containing 0.02% NaN₃, and 7 x 10¹² UV-inactivated blocking phage (see below); the final retentate (60-80 µl) is stored at 4 °C. BDP biotinylated with the NHS-SS-Biotin reagent is linked to 30 biotin via a disulfide-containing chain.

UV-irradiated M13 phage are used for blocking any biotinylated BDP which fortuitously binds filamentous phage in general. M13mp8 (Messing and Vieira, Gene 19: 262-276 (1982), which is incorporated herein by reference) is 5 chosen because it carries two amber mutations, which ensure that the few phage surviving irradiation will not grow in the sup O strains used to titer the surface expression library. A 5 ml sample containing 5×10^{13} M13mp8 phage, purified as described above, is placed in a small petri 10 plate and irradiated with a germicidal lamp at a distance of two feet for 7 minutes (flux 150 $\mu\text{W}/\text{cm}^2$). NaN₃ is added to 0.02% and phage particles concentrated to 10^{14} particles/ml on a Centricon 30-kDa ultrafilter (Amicon).

For panning, polystyrene petri plates (60 x 15 mm) are 15 incubated with 1 ml of 1 mg/ml of streptavidin (BRL) in 0.1 M NaHCO₃, pH 8.6-0.02% NaN₃, in a small, air-tight plastic box overnight in a cold room. The next day streptavidin is removed and replaced with at least 10 ml blocking solution (29 mg/ml of BSA; 3 $\mu\text{g}/\text{ml}$ of streptavidin; 0.1 M NaHCO₃, pH 20 8.6-0.02% NaN₃) and incubated at least 1 hour at room temperature. The blocking solution is removed and plates are washed rapidly three times with Tris buffered saline containing 0.5% Tween 20 (TBS-0.5% Tween 20).

Selection of phage expressing antibody fragments which 25 bind BDP is performed with 5 μl (2.7 μg BDP) of blocked biotinylated BDP reacted with a 50 μl portion of the library. Each mixture is incubated overnight at 4°C, diluted with 1 ml TBS-0.5% Tween 20, and transferred to a streptavidin-coated petri plate prepared as described 30 above. After rocking 10 minutes at room temperature, unbound phage are removed and plates washed ten times with TBS-0.5% Tween 20 over a period of 30-90 minutes. Bound phage are eluted from plates with 800 μl sterile elution buffer (1 mg/ml BSA, 0.1 M HCl, pH adjusted to 2.2 with 35 glycerol) for 15 minutes and eluates neutralized with 48 μl

2 M Tris (pH unadjusted). A 20 μ l portion of each eluate is titered on MK30-3 concentrated cells with dilutions of input phage.

A second round of panning is performed by treating 750 μ l of first eluate from the library with 5 mM DTT for 10 minutes to break disulfide bonds linking biotin groups to residual biotinylated binding proteins. The treated eluate is concentrated on a Centricon 30 ultrafilter (Amicon), washed three times with TBS-0.5% Tween 20, and concentrated to a final volume of about 50 μ l. Final retentate is transferred to a tube containing 5.0 μ l (2.7 μ g BDP) blocked biotinylated BDP and incubated overnight. The solution is diluted with 1 ml TBS-0.5% Tween 20, panned, and eluted as described above on fresh streptavidin-coated petri plates. The entire second eluate (800 μ l) is neutralized with 48 μ l 2 M Tris, and 20 μ l is titered simultaneously with the first eluate and dilutions of the input phage. If necessary, further rounds of panning can be performed to obtain homogeneous populations of phage. Additionally, phage can be plaque purified if reagents are available for detection.

Template Preparation and Sequencing

Templates are prepared for sequencing by inoculating a 1 ml culture of 2XYT containing a 1:100 dilution of an overnight culture of XL1 with an individual plaque from the purified population. The plaques are picked using a sterile toothpick. The culture is incubated at 37°C for 5-6 hours with shaking and then transferred to a 1.5 ml microfuge tube. 200 μ l of PEG solution is added, followed by vortexing and placed on ice for 10 minutes. The phage precipitate is recovered by centrifugation in a microfuge at 12,000 x g for 5 minutes. The supernatant is discarded and the pellet is resuspended in 230 μ l of TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) by gently pipeting with a yellow

5 pipet tip. Phenol (200 μ l) is added, followed by a brief vortex and microfuged to separate the phases. The aqueous phase is transferred to a separate tube and extracted with 200 μ l of phenol/chloroform (1:1) as described above for the phenol extraction. A 0.1 volume of 3 M NaOAc is added, followed by addition of 2.5 volumes of ethanol and precipitated at -20°C for 20 minutes. The precipitated templates are recovered by centrifugation in a microfuge at 12,000 \times g for 8 minutes. The pellet is washed in 70% 10 ethanol, dried and resuspended in 25 μ l TE. Sequencing was performed using a Sequenase™ sequencing kit following the protocol supplied by the manufacturer (U.S. Biochemical, Cleveland, OH).

EXAMPLE II

15 Cloning of Heavy and Light Chain Sequences
Without Restriction Enzyme Digestion

20 This example shows the simultaneous incorporation of antibody heavy and light chain fragment encoding sequences into a M13IXHL-type vector with the use of restriction endonucleases.

25 For the simultaneous incorporation of heavy and light chain encoding sequences into a single coexpression vector, a M13IXHL vector was produced that contained heavy and light chain encoding sequences for a mouse monoclonal antibody (DAN-18H4; Biosite, San Diego, CA). The inserted antibody fragment sequences are used as complementary sequences for the hybridization and incorporation of Hc and Lc sequences by site-directed mutagenesis. The genes encoding the heavy and light chain polypeptides were 30 inserted into M13IX30 (SEQ ID NO: 1) and M13IX11 (SEQ ID NO: 2), respectively, and combined into a single surface expression vector as described in Example I. The resultant M13IXHL-type vector is termed M13IX50.

The combinations were performed under conditions that facilitate the formation of one Hc and one Lc vector half into a single circularized vector. Briefly, the overhangs generated between the pairs of restriction sites after 5 restriction with Mlu I or Hind III and exonuclease digestion are unequal (i.e., 64 nucleotides compared to 32 nucleotides). These unequal lengths result in differential hybridization temperatures for specific annealing of the complementary ends from each vector. The specific 10 hybridization of each end of each vector half was accomplished by first annealing at 65°C in a small volume (about 100 µg/µl) to form a dimer of one Hc vector half and one Lc vector half. The dimers were circularized by diluting the mixture (to about 20 µg/µl) and lowering the 15 temperature to about 25-37°C to allow annealing. T4 ligas was present to covalently close the circular vectors.

M13IX50 was modified such that it did not produce a functional polypeptide for the DAN monoclonal antibody. To do this, about eight amino acids were changed within the 20 variable region of each chain by mutagenesis. The Lc variable region was mutagenized using the oligonucleotide 5'-CTGAACCTGTCTGGGACCACAGTTGATGCTATAGGATCAGATCTAGAATTCAATTAGAGACTGGCCTGGCTTCTGC-3' (SEQ ID NO: 68). The Hc sequence was mutagenized with the oligonucleotide 5'- 25 T C G A C C G T T G G T A G G A A T A A T G C A A T T A A T G; GAGTAGCTCTAAATTCTAGAATTCTACACCCAGTGCATCCAGTAGCT-3' (SEQ ID NO: 69). An additional mutation was also introduced into M13IX50 to yield the final form of the vector. During construction of an intermediate to M13IX50 (M13IX04 30 described in Example I), a six nucleotide sequence was duplicated in oligonucleotide 027 and its complement 032. This sequence, 5'TTACCG-3' was deleted by mutagenesis using the oligonucleotide 5'-GGTAAACAGTAACGGTAAGAGTGCCAG-3' (SEQ ID NO: 70). The resultant vector was designated M13IX53.

35 M13IX53 can be produced as a single stranded form and

contains all the functional elements of the previously described M13IXHL vector except that it does not express functional antibody heteromers. The single-stranded vector can be hybridized to populations of single-stranded Hc and 5 Lc encoding sequences for their incorporation into the vector by mutagenesis. Populations of single-stranded Hc and Lc encoding sequences can be produced by one skilled in the art from the PCR products described in Example I or by other methods known to one skilled in the art using the 10 primers and teachings described therein. The resultant vectors with Hc and Lc encoding sequences randomly incorporated are propagated and screened for desired binding specificities as described in Example I.

Other vectors similar to M13IX53 and the vectors it's 15 derived from, M13IX11 and M13IX30, have also been produced for the incorporation of Hc and Lc encoding sequences without restriction. In contrast to M13IX53, these vectors contain human antibody sequences for the efficient hybridization and incorporation of populations of human Hc 20 and Lc sequences. These vectors are briefly described below. The starting vectors were either the Hc vector (M13IX30) or the Lc vector (M13IX11) previously described.

M13IX32 was generated from M13IX30 by removing the six 25 nucleotide redundant sequence 5'-TTACCG-3' described above and mutation of the leader sequence to increase secretion of the product. The oligonucleotide used to remove the redundant sequence is the same as that given above. The mutation in the leader sequence was generated using the oligonucleotide 5'GGGCTTTGCCACAGGGT-3'. This mutagenesis 30 resulted in the A residue at position 6353 of M13IX30 being changed to a G residue.

A decapeptide tag for affinity purification of antibody fragments was incorporated in the proper reading frame at the carboxy-terminal end of the Hc expression site

in M13IX32. The oligonucleotide used for this mutagenesis was 5'-CGCCTT CAGCCTAAGAACGCTAGTCCGGAACGTCGTACGGTAGGATCCA CTAG-3' (SEQ ID NO: 71). The resultant vector was designated M13IX33. Modifications to this or other vectors 5 are envisioned which include various features known to one skilled in the art. For example, a peptidase cleavage site can be incorporated following the decapeptide tag which allows the antibody to be cleaved from the gene VIII portion of the fusion protein.

10 M13IX34 (SEQ ID NO: 3) was created from M13IX33 by cloning in the gene encoding a human IgG1 heavy chain. The reading frame of the variable region was changed and a stop codon was introduced to ensure that a functional polypeptide would not be produced. The oligonucleotide 15 used for the mutagenesis of the variable region was 5'-CACCGGTTGGGGATTAGTCTTGACCAAGGCAGGCCAGGGC-3' (SEQ ID NO: 72). The complete nucleotide sequence of this vector is shown in Figure 4 (SEQ ID NO: 3).

20 Several vectors of the M13IX11 series were also generated to contain similar modifications as that described for the vectors M13IX53 and M13IX34. The promoter region in M13IX11 was mutated to conform to the 35 consensus sequence to generate M13IX12. The oligonucleotide used for this mutagenesis was 5'-ATTCCACAC 25 ATTATACGAGCCGGAAGCATAAAAGTGTCAAGCCTGGGTGCC-3' (SEQ ID NO: 73). A human kappa light chain sequence was cloned into M13IX12 and the variable region subsequently deleted to generate M13IX13 (SEQ ID NO: 4). The complete nucleotide sequence of this vector is shown in Figure 5 (SEQ ID NO: 30 4). A similar vector, designated M13IX14, was also generated in which the human lambda light chain was inserted into M13IX12 followed by deletion of the variable region. The oligonucleotides used for the variable region deletion of M13IX13 and M13IX14 were 5'-CTG 35 CTCATCAGATGGCGGGAAAGAGCTCGGCCATGGCTGGTTG-3' (SEQ ID NO: 74)

and 5'-GAACAGAGT GACCGAGGGGGCGAGCTCGGCCATGGCTGGTTG-3' (SEQ ID NO: 75), respectively.

The Hc and Lc vectors or modified forms thereof can be combined using the methods described in Example I to 5 produce a single vector similar to M13IX53 that allows the efficient incorporation of human Hc and Lc encoding sequences by mutagenesis. An example of such a vector is the combination of M13IX13 with M13IX34. The complete 10 nucleotide sequence of this vector, M13IX60, is shown in Figure 6 (SEQ ID NO: 5).

Additional modifications to any of the previously described vectors can also be performed to generate vectors which allow the efficient incorporation and surface expression of Hc and Lc sequences. For example, to 15 alleviate the use of uracil selection against wild-type template during mutagenesis procedures, the variable region locations within the vectors can be substituted by a set of palindromic restriction enzyme sites (i.e., two similar sites in opposite orientation). The palindromic sites will 20 loop out and hybridize together during the mutagenesis and thus form a double-stranded substrate for restriction endonuclease digestion. Cleavage of the site results in the destruction of the wild-type template. The variable region of the inserted Hc or Lc sequences will not be 25 affected since they will be in single stranded form.

Following the methods of Example I, single-stranded Hc or Lc populations can be produced by a variety of methods known to one skilled in the art. For example, the PCR primers described in Example I can be used in asymmetric 30 PCR to generate such populations. Gelfand et al., "PCR Protocols: A Guide to Methods and Applications", Ed by M.A. Innis (1990), which is incorporated herein by reference. Asymmetric PCR is a PCR method that differentially amplifies only a single strand of the double

stranded template. Such differential amplification is accomplished by decreasing the primer amount for the undesirable strand about 10-fold compared to that for the desirable strand. Alternatively, single-stranded 5 populations can be produced from double-stranded PCR products generated as described in Example I except that the primer(s) used to generate the undesirable strand of the double-stranded products is first phosphorylated at its 5' end with a kinase. The resultant products can then be 10 treated with a 5' to 3' exonuclease, such as lambda exonuclease (BRL, Bethesda, MD) to digest away the unwanted strand.

Single-stranded Hc and Lc populations generated by the methods described above or by others known to one skilled 15 in the art are hybridized to complementary sequences encoded in the previously described vectors. The population of the sequences are subsequently incorporated into a double-stranded form of the vector by polymerase extension of the hybridized templates. Propagation and 20 surface expression of the randomly combined Hc and Lc sequences are performed as described in Example I.

Although the invention has been described with reference to the presently preferred embodiment, it should be understood that various modifications can be made 25 without departing from the spirit of the invention. Accordingly, the invention is limited only by the claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: HUSE, WILLIAM D.

(ii) TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF HETEROMERIC RECEPTORS

(iii) NUMBER OF SEQUENCES: 75

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: PRETTY, SCHROEDER, BRUEGEMANN & CLARK
 (B) STREET: 444 SO. FLOWER STREET, SUITE 200
 (C) CITY: LOS ANGELES
 (D) STATE: CALIFORNIA
 (E) COUNTRY: UNITED STATES
 (F) ZIP: 90071

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
 (B) FILING DATE:
 (C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: CAMPBELL, CATHRYN A.
 (B) REGISTRATION NUMBER: 31,815
 (C) REFERENCE/DOCKET NUMBER: P31 8882

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 619-535-9001
 (B) TELEFAX: 619-535-8949

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7445 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAAG CTCGGCCCCC AAATGAAAT	60
ATAGCTAAC AGGTTATTGA CCATTTGCCA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCGAGA ATTGGGAATC AACTGTTACA TCGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTG AGCAATTAAG CTCTAAGGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTG CTTCCGGTCT GGTTCCGCTT GAAGCTCGAA TTAAAACCCG ATATTTGAAG	360
TCTTTGGGCC TTCCCTTAA TCTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT	420

CAGGGTAAAG ACCTGATTT TGATTTATGG TCATTCTCGT TTTCTGAAC	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACCC	540
TATCCAGTCT AAACATTTA CTATTACCCC CTCTGGCAAA ACTTCTTTG CAAAAGCCTC	600
TCGCTATTT GGTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG	660
TTGCTCTTAC TATGCCCTCGT AATTCCCTTT GGCGTTATGT ATCTGCATTA	720
GTTGAATGTG GTATTCCCAA ATCTCAACTG ATGAATCTTT CTACCTGTAA	780
TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTT TCTTCCCAAAC	840
GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA	900
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT	960
TCTGGTGTTC CTCGTCAAGG CAAGCCTTAT TCAGTGAATG AGCAGCTTG	1020
TTACGTTGAT TTGGGTAATG AATATCCGGT TCTTGTCAAG ATTACTCTTG	1080
ATGAAGGTCA GCCAGCCTAT CGCCCTGGTC TGTACACCGT TCATCTGTCC	1140
TCTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC GTCTGCCCT	1200
CGTTCCGGCT AAGTAACATG GAGCAGGTGCG CGGATTTCGA CACAATTAT	1260
CAGGCCATGA TACAAATCTC CGTTGTACTT TGTTTCCGGC TTGGTATAAT	1320
CGCTGGGGGT CAAAGATGAG TGTTTGTAGTG TATTCTTCG CCTCTTCTGT	1380
TTAGGTTGG TGCCCTTCGTA GTGGCATTAC GTATTTACC CGTTTAATGG	1440
AAACTTCCTC ATGAAAAAGT CTTTAGTCCT CAAAGCCTCT GTAGCCGTTG	1500
CTACCCCTCGT TCCGATGCTG TCTTCTGCTG CTGAGGGTGA CGATCCCGCA	1560
AAAGCCGCCT TTAACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1620
TGGTGGGGCG ATGGTTGTTG TCATTGTCGG CGCAACTATC GGTATCAAGC	1680
TGTTTAAGAA ATTACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA	1740
GGCTCCTTTT GGAGCCTTTT TTTTGGAGA TTTTCAACGT GAAAAAATTA	1800
TTATTCGCAA TTCCCTTAGT TGTTCCCTTC TATTCTCACT CCCGTAAAC	1860
TGTTGAAAGT TGTTAGCAA AACCCCATAC AGAAAATTCA TTACTAACG	1920
TCTGGAAAGA CGACAAAAGT TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1980
CTGTGGAATG CTACAGGCCTG TGTAGTTGT ACTGGTGACG AAACTCAGTG	2040
TTACGGTACA TGGGTTCTTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG	2100
GTGGCTCTGA GGTTGGCGGT ACTAACCTC CTGAGTACGG TGATACACCT	2160
ATTCCGGGCT ATACTTATAT CAACCCCTCTC GACGGCACTT ATCCGCCTGG	2220
TACTGAGCAA AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC	2280
CTCTTAATAC TTTCATGTT CAGAATAATA GGTTCCGAAA TAGGCAGGGG	2340
GCATTAACIG TTTATACGGG CACTGTTACT CAGTACACTC CTGTATCATC	2400
AAAAGCCATG TATGACCGCT ACTGGAACGG TAAATTAGA GACTGCGCTT	2460
TCCATTCTGG CTTTAATGAA GATCCATTCTG TTTGTGAATA TCAAGGCCAA	
TCGTCTGACC TGCCTCAACC TCCTGTCAAT GCTGGGGGG CCTCTGGTGG	
TGGTTCTGGT GGCGGCTCTG AGGGTGGTGG CTCTGAGGGT GGCGGTTCTG	
AGGGTGGCGG CTCTGAGGGA GGCGGTTCCG GTGGTGGCTC TGGTTCCGGT	
GATTTTGATT ATGAAAAGAT GGCAAAACGCT AATAAGGGGG CTATGACCGA	
AAATGCCGAT	

GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACTTGATT CTGTCGCTAC TGATTACCGT	2520
GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTACACCT	2640
TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTTGA ATGTCCCCCT	2700
TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTTGTGACAA AATAAACTTA	2760
TTCCGTGGTG TCTTTCGCGT TCTTTTATAT GTTGCACCT TTATGTATGT ATTTTCTACCG	2820
TTTGTCTAACAA TACTGCCCTAA TAAGGAGTCT TAATCATGCC AGTTCTTTTG GGTATTCCGT	2880
TATTATTGCG TTTCTCGGT TTCCCTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTTCCGTAAG ATAGCTATTG CTATTTCAATT GTTCTTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTTG TGTTATCTCT CTGATATTAG CGCTCAATTAA CCCTCTGACT	3060
TTGTTCAAGGG TGTTCACTTA ATTCTCCGT CTAATGCCCT TCCCTGTTTT TATGTTATTG	3120
TCTCTGTAAA GGCTGCTATT TTCACTTTTG ACGTTAAACA AAAAATCGTT TCTTATTG	3180
ATTGGGATAAA ATAATATGGC TGTTTATTTT GTAACTGGCA AATTAGGCTC TGGAAAGACGG	3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT	3300
CTTGATTAA GGCTTCAAAA CCTCCCCCAA GTCGGGAGGT TCGCTAAAAC CCCTCCGGTT	3360
CTTAGAATAC CCGATAAGCC TTCTATATCT GATTTGCTTG CTATTGGCG CGGTAATGAT	3420
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGCCTGAC TTGGTTAAT	3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT TTTCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGGC	3600
CGTTCTGCAT TAGCTGAACA TGTTGTTAT TGTCTCGTC TGGACAGAAT TACTTACCT	3660
TTTGTGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
GTGCGGTTG TAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTAT	3780
ACTGGTAAGA ATTGTATAAA CCCATATGAT ACTAACACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTGTT ATTCTTATTAA AACGCCCTAT TTATCACACCG GTCGGTATT CAAACCATTA	3900
AATTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGAA AAAAGTTTC ACGCGTTCTT	3960
TGTCTTGCGA TTGGATTTCG ATCAGCATT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGACTCTCT	4080
CACCGCTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGAAA ATTAATTAAT	4140
AGCCACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTATG TACTGTTCC	4200
ATTAAAAAG GTAATTCAA TGAAATTGTT AAATGTAATT AATTTGTTT TCTTGTATGTT	4260
TGTTTCATCA TCTTCTTTTG CTCAGGTAAT TGAAATGAAT AATTGCCCTC TGGCGATT	4320
TGTAACCTGG TATTCAAAGC AATCAGGCCA ATCCGTTATT GTTCTCCCG ATGTAAAAGG	4380
TACTGTTACT GTATATTCACT CGACCTTAA ACCTGAAAAT CTACCGAATT TCTTATTTC	4440
TGTTTACGT GCTAATAATT TTGATATGGT TGCTCAATT CCTTCCATAA TTCAAGAGTA	4500

TAATCCAAAC AATCAGGATT ATATTGATCA ATTGCCATCA TCTGATAATC AGGAATATGA 4560
 TGATAATTCC GCTCCTTCTG GTGGTTTCTT TGTTCGGCAA AATGATAATG TTACTCAAAC 4620
 TTTTAAATT AATAACGTTG GGGCAAAGGA TTTAATACGA GTTGTGAAAT TGTTTGTAAA 4680
 GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT 4740
 TAGTGCACCT AAAGATATTT TAGATAACCT TCCTCAATTC CTTTCTACTG TTGATTGCGC 4800
 AACTGACCAAG ATATTGATTG AGGGTTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTAGA 4860
 TTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG 4920
 CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTGCTTCGGT ATTTTAAATG GCGATGTTTT 4980
 AGGGCTATCA GTTCGGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG 5040
 TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GCCCAGAATG TCCCTTTAT 5100
 TACTGGTCTG GTGACTGGTG AATCTGCCAA TGTAAATAAT CCATTTCAGA CGATTGAGCG 5160
 TCAAAATGTA CGTATTTCCA TGAGCGTTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT 5220
 TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCAA GTGATGTTAT 5280
 TACTAATCAA AGAAGTATTG CTACAACGGT TAATTGCGT GATGGACAGA CTCTTTACT 5340
 CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCTGTCIAA 5400
 AATCCCTTTA ATCGGCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGGG AAAGCACGTT 5460
 ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCCATTAA AGCCGGCCGG 5520
 GTGTGGTGGT TACGGCGAGC GTGACCGCTA CACTGCCAG CGCCCTAGCG CCCGCTCCTT 5580
 TCGCTTTCTT CCCTTCCCTT CTGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC 5640
 GGGGGCTCCC TTIAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCCC AAAAAACTTG 5700
 ATTTGGGTGA TGGTTCACGT ACTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA 5760
 CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTCCA AACTGGAACA ACACCTCAACC 5820
 CTATCTCGGG CTATTCTTT GATTATAAG GGATTTGCG GATTTCGGAA CCACCATCAA 5880
 ACAGGATTTC CGCCTGCTGG GGCAAACCAAG CGTGGACCGC TTGCTGCAAC TCTCTCAGGG 5940
 CCAGGCGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCT 6000
 GGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTGGCC GATTCAATTAA TGCAGCTGGC 6060
 ACCACAGGTT TCCCAGTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAAT GTGAGTTAGC 6120
 TCACTCATTAA GGCAACCCAG GCTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGAA 6180
 TTGTGAGCGG ATAACAATT T CACACGGTC ACTTGGCACT GGCGTCCGT TTACAACGTC 6240
 GTGACTGGGA AAACCCCTGGC GTTACCCAAG CTITGTACAT GGAGAAAATA AAGTGAACA 6300
 AAGCACTATT GCACTGGCAC TCTTACCGTT ACCGTTACTG TTTACCCCTG TGACAAAAGC 6360
 CGCCCGAGGTC CAGCTGCTCG ACTCAGGGCT ATTGTCCCCA GGGGATTGTA CTAGTGGATC 6420
 CTAGGCTGAA GGCGATGACC CTGCTAAGGC TGCATTCAAT AGTTTACAGG CAACTGCTAC 6480
 TGAGTACATT GGCTACGGCTT CGGCTATGGT ACTAGTTATA GTTGGTGTCA CCATAGGGAT 6540

TAAATTATTTC AAAAAGTTTA CGACCAAGGC TTCTTAAGCA ATAGCGAAGA GGCCCCGACC 6600
 GATCGCCCTT CCCAACAGTT CGCGACCCCTG AATGGCGAAT GGCGCTTGC CTGGTTCCG 6660
 GCACCCAGAAG CGGTGCCCGA AAGCTGGCTG GAGTGGCATC TTCCCTGAGGC CGATACGGTC 6720
 GTCGTCCCTT CAAACTGGCA GATCCACGGT TACCGATGCCG CCATCTACAC CAACGTAACC 6780
 TATCCCATTA CGGTCAATCC GCCGTTTGTG CCCACGGAGA ATCCGACGGG TTGTTACTCG 6840
 CTCACATTTA ATGTTGATCA AAGCTGGCTA CAGGAAGGCC AGACCGAAT TATTTTGAT 6900
 CGCGTTCCCTA TTGGTTAAAA AATGAGCTGA TTAAACAAAA ATTTAACCGG AATTTTAACA 6960
 AAATATTAAC GTTACAAATT TAAATATTTG CTTATACAAT CTTCCCTGTT TTGGGGCTTT 7020
 TCTGATTATC AACCAGGGTA CATATGATTG ACATGCTAGT TTACGATTA CGGTTCATCG 7080
 ATTCTCTTGT TTGCTCCAGA CTCTCAGGCA ATGACCTGAT AGCCTTGTG GATCTCTCAA 7140
 AAATAGCTAC CCTCTCCGGC ATTAATTAT CAGCTAGAAC CGTTGAATAT CATATTGATG 7200
 GTGATTTGAC TGTCTCCGGC CTTTCTCACC CTTTGAATC TTTACCTACA CATTACTCAG 7260
 GCATTGCATT TAAAATATAT GAGGGTCTA AAAATTTTA TCCTTGCCTT GAAATAAAGG 7320
 CTTCTCCGGC AAAAGTATTA CAGGGTCATA ATGTTTTGG TACAACCGAT TTAGCTTTAT 7380
 GCTCTGAGGC TTATGCTT AATTTGCTA ATTCTTGCC TTGCCTGTAT GATTTATTGG 7440
 ACGTT 7445

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7317 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTCTAG CTCGGCCCCC AAATGAAAAT 60
 ATAGCTAAAC AGGTATTGA CCATTTGCCA AATGTATCTA ATGGTCAAAC TAAATCTACT 120
 CGTTCCAGA ATTGGGAATC AACTGTTACA TGGAAATGAAA CTTCCAGACA CCGTACTTTA 180
 GTTCCATATT TAAAACATGT TGAGCTACAG CACCAAGATT AGCAATTAAAG CTCTAAGCCA 240
 TCCGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG 300
 TTGGAGTTTG CTTCCGGTCT GGTTCCCTT GAAGCTCGAA TTAAAACCGG ATATTTGAAG 360
 TCTTTGGGC TTCCCTTTAA TCTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT 420
 CAGGGTAAAG ACCTGATTT TGATTTATGG TCATTCTCGT TTTCTGAACG GTTTAAAGCA 480
 TTTCAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT 540
 AAACATTTTA CTATTACCCC CTCTGGAAA ACTTCTTTTG CAAAAGCCTC TCGCTATTTT 600
 CGTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT 660
 AATTCCCTTTT GGCGTTATGT ATCTGCATTA GTTGAATGTC GTATTCCCTAA ATCTCAACTG 720

ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATTT	780
TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCCATA	AGGTAATTCA	840
CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGT	900
CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTC	TTACGTTGAT	TTGGGTAATG	960
AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC	1020
TGTACACCGT	TCATCTGTCC	TCTTCAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC	1080
GTCTGCCCT	CGTCCGGCT	AACTAACATG	GACCAGGTG	CGGATTTCGA	CACAAATTAT	1140
CAGGGCATGA	TACAAATCTC	CGTTGTACTT	TGTTTCCGGC	TTGGTATAAT	CGCTGGGGT	1200
CAAAGATGAG	TGTTTTAGTG	TATTCTTTCG	CCTCTTTCGT	TTAGGTTGG	TGCCTCGTA	1260
GTGGCATTAC	GTATTTACC	CGTTTAATGG	AAACTTCCTC	ATGAAAAAGT	CTTTAGTCCT	1320
CAAACCTCT	GTAGCCGTG	CTACCCCTCGT	TCCGATGCTG	TCTTTCGCTG	CTGAGGGTGA	1380
CGATCCCGCA	AAAGCCGGCT	TTAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	1440
TGCGTGGGGC	ATGGTTGTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA	1500
ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCCTTT	GGAGCCTTTT	1560
TTTTGGAGA	TTTCAACGT	GAAAAAAITA	TTATTCCCAA	TTCTTTAGT	TGTTCCCTTC	1620
TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA	1680
TTTACTAACG	TCTGGAAAGA	CGACAAAACT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
CTGTGGAATG	CTACAGGCCT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA	1800
TGGGTTCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860
TCTGAGGGTG	GGGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGACTACGG	TGATACACCT	1920
ATTCCGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACTT	ATCCGCCTGG	TACTGAGCAA	1980
AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT	2040
CAGAATAATA	GGTCCGAAA	TAGGCAGGGG	GCATTAACTG	TTTATACGGG	CACTGTTACT	2100
CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG	2160
TATGACCGCTT	ACTGGAACGG	TAAATTCAAG	GAATGCCCTT	TCCATTCTGG	CTTTAATGAA	2220
GATCCATTCC	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCCTCAACC	TCCGTCAAT	2280
GCTGGCGGGC	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340
GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT	2400
GATTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	2460
AAAAACGCCG	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	2520
GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT	2580
GGTGATTTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCACCT	2640
TTAATGAATA	ATTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT	2700
TTTGTCTTTA	GGCGTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTACCAA	AATAAACTTA	2769

TTCCGTGGTG TCTTTCGGTT TCTTTATAT GTGCCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAACAA TACTGGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTG GGTATTCCGT	2880
TATTATTGCG TTTCTCGGT TTCTCTGG TAACCTTGTG CGGCTATCTG CTTACTTTC	2940
TTAAAAAGGG CTTCGTAAG ATAGCTATTG CTATTCATT GTTCTTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTTGTG GTTATCTCT CGATATTAG CGCTCAATTAA CCCTCTGACT	3060
TTGTTCAAGGG TGTTCAAGTTA ATTCTCCCGT CTAATGCCGT TCCCTGTTT TATGTTATTG	3120
TCTCTGTAAA GGCTGCTATT TTCACTTTTG ACGTTAAACA AAAAATCGTT TCTTATTG	3180
ATTGGGATAAA ATAATATGGC TGTTTATTT GTAACTGGCA AATTAGGCTC TGGAAAGACG	3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGAAAAT AGCAACTAAT	3300
CTTGATTAA GGCTCAAAA CCTCCCGAA GTCGGGAGGT TCGCTAAAC GCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTTGCTTG CTATTGGCG CGGTAATGAT	3420
TCCTACGATG AAAATAAAA CGGCTTGCTT GTTCTCGATG AGTGGGGTAC TTGGTTAAAT	3480
ACCCGTTCTT GGAATGATAA CGAAAGACAG CCGATTATTG ATTGGTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT TTTCCTTGTG CAGGACTTAT CTATTGTTGA TAAACAGGGC	3600
CGTTCTGCAT TAGCTGAACA TGTTGTTAT TGTCGTCGTC TGGACAGAAAT TACTTTACCT	3660
TTTGTGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGGG TTGGCTTTAT	3780
ACTGGTAAGA ATTGTATAA CGCATATGAT ACTAACACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTGTT ATTCTTATTT AACGCCCTAT TTATCACACG GTCGGTATTT CAAACCATTAA	3900
AATTAGGTC AGAACATGAA CCTTACTAAA ATATATTGA AAAAGTTTTC ACCCGTTCTT	3960
TGTCTTGGCA TTGGATTTCG ATCAGCATT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTAAAAA AGTAGTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGACTCTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAAT	4140
AGCCACGATT TACAGAAGCA AGGTATTCA CTCACATATA TTGATTATG TACTGTTCC	4200
ATTAAAAAAG GTAATTCAA TGAAATTGTT AAATGTAATT AATTGTTTT TCTTGATGTT	4260
TGTTTCATCA TCTTCTTTG CTCAGGTAAT TGAAATGAAT AATTGCCCTC TGCGGGATT	4320
TGTAACCTGG TATTCAAAGC AATCAGGGCA ATCCGTTATT GTTCTCCCG ATGAAAAGG	4380
TACTGTTACT GTATATTCACT CTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTATTC	4440
TGTTTACGT GCTAATAATT TTGATATGGT TGTTCAATT CCTTCCATAA TTCAGAAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATCA	4560
TGATAATTCC GCTCCTCTG GTGGTTCTT TGTTCCCAA AATGATAATC TTACTCAAAC	4620
TTTTAAAATT AATAACGTTG GGGCAAAGGA TTAAATACGA GTTGTGAAT TGTTGTAAGA	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT	4740
TAGTGACCT AAAGATATTG TAGATAACCT TCCTCAATTC CTTTCTACTG TTGATTGCC	4800

AACTCACCAAG ATATTGATTG AGGGTTTGT ATTTGAGGTT CAGCAAGGTG ATGCCTTACA 4860
 TTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGGCGTGTAA ATACTGACCG 4920
 CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTGCGTTCGGT ATTTTTAATG GCGATGTGTT 4980
 AGGCCTATCA GTTCCGCCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG 5040
 TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GCCCAGAATG TCCCTTTAT 5100
 TACTGCTCGT GTGACTGGTG AATCTGCCAA TGTAAATAAT CCATTTAGA CGATTGAGCG 5160
 TCAAAATGTA CGTATTTCCA TGAGCGTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT 5220
 TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCAA GTGATGTTAT 5280
 TACTAATCAA AGAAGTATTG CTACAAACGGT TAATTGCGT GATGGACAGA CTCTTTACT 5340
 CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCGTCTAA 5400
 AATCCCTTTA ATCGGCCTCC TGTTAGCTC CCGCTCTGAT TCCAACGGAG AAAGCACGTT 5460
 ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGGCCATTA AGCGCGGGCG 5520
 GTGTGGTGGT TACGCCAGC GTGACCGCTA CACTGCCAG CGCCCTAGCG CCCGCTCCTT 5580
 TCGCTTTCTT CCCTTCCTTT CTGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC 5640
 GGGGGCTCCC TTTAGGGTTT CGATTTAGTG CTTTACGGCA CCTCGACCCCC AAAAAAACTTG 5700
 ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA 5760
 CGTTGGAGTC CACGTTCTT AATAGTGGAC TCTTGTCCA AACTGGAACA ACACCTAAC 5820
 CTATCTCGGG CTATTCTTT GATTATAAG GGATTTTGCC GATTTGGAA CCACCATCAA 5880
 ACAGGATTTC CGCCTGCTGG GGCAAACCCAG CGTGGACCCG TTGCTGCAAC TCTCTCAGGG 5940
 CCAGGCGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCT 6000
 GGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCACTAA TGCAGCTGGC 6060
 ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC 6120
 TCACTCATTAA GGCAACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA 6180
 TTGTGAGCGG ATAACAATT TACACGCCAA GGAGACAGTC ATAATGAAAT ACCTATTGCC 6240
 TACGGCAGCC GCTGGATTGT TATTACTCGC TCCCCAACCA GCCATGGCCG AGCTCGTGAT 6300
 GACCCAGACT CCAGATATCC AACAGGAATG AGTGTAAATT CTAGAACCGG TCACCTGGCA 6360
 CTGGCCGTGG TTTTACAACG TCGTGAATGG 6420
 GAAAACCTG GCGTTACCCA AGCTTAATCG
 CCTTGCAGAA TTCCCTTTCG CCAGCTGGGG TAATAGCGAA GAGGCCCGCA CCGATCCCC 6480
 TTCCCAACAG TTGCCAGCC TGAATGGCGA ATGGCCCTT GCCTGGTTTC CGGCACCAGA 6540
 AGCGGTGCCG GAAAGCTGGC TCGAGTGCAG TCTTCTGAG GCCGATACGG TCGTCGTCCC 6600
 CTCAAACCTGG CAGATGCACG GTTACCGATGC GCCCATCTAC ACCAACCTAA CCTATCCCCT 6660
 TACGGTCAAT CCGCCGTGG TTCCCAACCGA GAATCCGAGG GGTTGTTACT CGCTCACATT 6720
 TAATGTTGAT CAAACCTGGC TACAGGAAGG CCAGACCCGA ATTATTTTG ATGGCCTTCC 6780
 TATTGCTTAA AAAATGAGCT GATTTAACAA AAATTTAACG CGAATTTAA CAACAAATATA

ACGTTTACAA TTTAAATATT TGCTTATACA ATCTTCCTGT TTTGGGGCT TTTCTGATTA 6900
 TCAACCGGGG TACATATGAT TGACATGCTA GTTTACCGAT TACCGTTCAT CGATTCTCTT 6960
 GTTTGCTCCA GACTCTCAGG CAATGACCTG ATAGCCTTTC TAGATCTCTC AAAAATAGCT 7020
 ACCCTCTCCG GCATTAATTI ATCAGCTAGA ACGGTTGAAT ATCATATTGA TGGTGATTTC 7080
 ACTCTCTCCG GCCTTCTCA CCCTTTGAA TCTTTACCTA CACATTACTC AGGCATTGCA 7140
 TTTAAAATAT ATGAGGGTTC TAAAAATTTC TATCCTTGGG TTGAAATAAA GGCTTCTCCC 7200
 GCAAAAGTAT TACAGGGTCA TAATGTTTTT GGTACAACCG ATTTAGCTTT ATGCTCTCAG 7260
 GCTTTATTGC TTAATTTCGC TAATTCTTTC CCTTGCTGT ATGATTTATT GGATGTT 7317

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7729 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTCAAG CTCGGGCCCC AAATGAAAAT 60
 ATAGCTAAAC AGGTTATTGA CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT 120
 CGTTCCGAGA ATTGGGAATC AACTGTTACA TCGAATGAAA CTTCCAGACA CCGTACTTTA 180
 GTTGCATATT TAAAACATGT TGAGCTACAG CACCAAGATTAGC AGCAATTAAAG CTCTAAGCCA 240
 TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG 300
 TTGGAGTTTG CTTCCGGTCT CGTTCCGTTT GAAGCTCGAA TTAAAACCGG ATATTTGAAC 360
 TCTTTGGGC TTCCCTTTAA TCTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT 420
 CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACG GTTTAAAGCA 480
 TTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGGAG TATTGGACCC TATCCAGTCT 540
 AAACATTITA CTATTACCCC CTCTGGAAA ACTTCTTTG CAAAAGCCTC TCGCTATTIT 600
 GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCCTCGT 660
 AATTCCCTTT GGCCTTATGT ATCTCCATTA GTTGAATGTC GTATTCTAA ATCTCAACTG 720
 ATGAATCTTT CTACCTGTAA TAATGTTGTT CGCTTAGTTG GTTTTATTAA CGTAGATTT 780
 TCTTCCCAAC GTCCGTACTG GTATAATGAG CCAGTTCTTA AAATCCATA AGGTAATTCA 840
 CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGT 900
 CTCGTCAAGG CAAGCCTTAT TCACTGAATG AGCAGCTTTC TTACGTTGAT TTGGGTAAATG 960
 AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC 1020
 TGTACACCGT TCATCTGTCC TCTTTCAAAG TTGGTCAGTT CGGTCCCTT ATGATTGACC 1080
 GTCTCCGGCT CGTTCCGGCT AAGTAACATG GAGCAGGTGCG CGGATTTCGA CACAATTAT 1140
 CAGGGGATGA TACAAATCTC CGTTGTACTT TGTGTTGGGGC TTGCTATAAT CGCTGGGGCT 1200

CAAAGATGAG TGTGTTAGTC TATTCCTTCG CCTCTTTCGT TTTAGGTTGG TGCCTTCGTA	1260
GTGGCATTAC CTATTTTACCG GTTTAATGG AAACCTCCTC ATGAAAAAGT CTTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTG CTACCCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA	1380
CGATCCCCA AAAGCGGCCT TTAACCTCCCT GCAAGCCTCA GCGACCGAAT ATATCCGTTA	1440
TGCCCTGGCC ATGCTTGTG TCATTTGTCGG CGCAACTATC GGTATCAACC TGTTTAAGAA	1500
ATTCAACCTCG AAACCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTT GGAGCCCTTT	1560
TTTTTGGAGA TTTTCAACGT GAAAAAAITA TTATTCGCAA TTCTTCTAGT TGTTCTTTC	1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTTACCAA AACCCCATAC AGAAAATTCA	1680
TTTACTAACG TCTGGAAAGA CGACAAAAGT TTAGATCGTT ACCGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGGCT TGTAGTTGT ACTGGTGACG AAACTCAGTG TTACGGTACA	1800
TGGCTTCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACCG TGATACACCT	1920
ATTCCGGGCT ATACTTATAT CAACCCCTCTC GACGGCACTT ATCCGGCTGG TACTGAGCAA	1980
AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT	2040
CAGAATAATA CGTTCCGAAA TAGGCAGGGG CGATTAACGT TTTATACGGG CACTGTACT	2100
CAAGGCACGT ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG	2160
TATGACGCTT ACTGGAACGG TAAATTAGA GACTGGCCTT TCCATTCTGG CTTTAATGAA	2220
GATCCATTCG TTGTAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC TCCTGTCAAT	2280
GCTGGCGGCG GCTCTGGTGG TGTTCTGGT GGCGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA GGCGGTTCCG GTGGTGGCTC TGTTCCGGT	2400
GATTTGATT ATGAAAAGAT GGCAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	2460
GAAAACGGCG TACAGTCTGA CGCTAAAGGC AAACCTGATT CTGTCGCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCAT TGTTGACGTT TCCGGCCTTG CTAATGGTAA TGTTGCTACT	2580
GGTGTATTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTACCT	2640
TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTGA ATGTCGCCCT	2700
TTTGTCTTAA CGCGTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAAACTTA	2760
TTCCGTGGTG TCTTTATAT GTGCCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAACAA TACTGGCTAA TAAGGAGTCT TAATCATGCC AGTTCTTTG GGTATTCCGGT	2880
TATTATTGCG TTCCCTCGGT TTCCCTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTTCCGTAAG ATAGCTATTG CTATTCATT GTTCTTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTTGCG GGTTATCTCT CTGATATTAG CGCTCAATTAA CCCTCTGACT	3060
TTGTTCAAGGG TGTTCAAGTAA ATTCTCCCGT CTAATGCCCT TCCCTGTTT TATGTTATTG	3120
TCTCTGTAAA CGCTGCTATT TTCAATTGACGTTAAACAA AAAAATCGTT TCTTATTGCG	3180
ATTGGGATAAA ATAATATGCC TGTTTATTTT CTAACGGCA AATTAGGCTG TGGAAGACG	324

CTCGTTAGCC TTGCTAAGAT TCAGGATAAA ATTGTAGCTG CGTGCAAAAT AGCAACTAAT 3300
 CTTGATTTAA GGCTTCAAAA CCTCCCCAA GTCGGGAGGT TCGCTAAAAC GCCTCGCGT 3360
 CTTAGAATAC CGCATAAGCC TTCTATATCT GATTGCTTG CTATTGGCG CCGTAATCAT 3420
 TCCTACCGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGCAGTAC TTGCTTAAT 3480
 ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATIG ATTGGTTCT ACATGCTCGT 3540
 AAATTAGGAT GGGATATTAT TTTTCTTGT CAGGACTTAT CTATTGTTGA TAAACAGGCG 3600
 CGTTCTGCAT TAGCTGAACA TGTGTTTAT TGTCTCGTC TGGACAGAAT TACTTTACCT 3660
 TTTGTCGGTA CTTTATATTTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT 3720
 GTTGGCGTTG TAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCC TTGGCTTAT 3780
 ACTGGTAAGA ATTGTATAA CGCATATGAT ACTAAACAGG CTTTTCTAG TAATTATCAT 3840
 TCCGGTGTGTT ATTCTTATTT AACGCCCTTAT TTATCACACG GTCCGTATTT CAAACCATTA 3900
 AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGA AAAAGTTTTC ACGCGTTCTT 3960
 TGTCTTGCCTA TTGGATTTCG ATCAGCATT ACATATAGTT ATATAACCCA ACCTAACCCG 4020
 GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTTGATA AATTCACTAT TGACTCTTCT 4080
 CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAAT 4140
 AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTCC 4200
 ATTAAAAAAG CTAATTCAA TGAAATTGTT AAATGTAATT AATTTCTTT TCTTGATGTT 4260
 TGTTTCATCA TCTTCTTTG CTCAGGTAAT TGAAATGAAT AATTGGCCTC TGGCGATT 4320
 TGTAACTTGG TATTCAAAGC AATCAGGCCA ATCCGTTATT GTTTCTCCCG ATGTAAGG 4380
 TACTGTTACT GTATATTCACT CTGACGTTAA ACCTGAAAAT CTACCGAATT TCTTCTTTC 4440
 TGTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCCATAA TTCAAGAGTA 4500
 TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA 4560
 TGATAATTCC GCTCCTCTG GTGGTTCTT TGTTCCGAA AATGATAATG TTACTCAAAC 4620
 TTTTAAATT AATAACGTTG GGGCAAACGA TTTAATACGA GTTGTGAAT TCTTGTAAA 4680
 GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT 4740
 TAGTGCACCT AAAGATATT TAGATAACCT TCCTCAATT CTTTCTACTG TTGATTTGCC 4800
 AACTGACCAAG ATATTGATTG AGGGTTGAT ATTGAGGTT CAGCAAGGTG ATGCTTCTA 4860
 TTTTCACTT GCTGCTGGCT CTCAGCGTGG CACTGTGCA GGCGGTGTTA ATACTGACCC 4920
 CCTCACCTCT GTTTATCTT CTGCTGGTGG TTGTTCCGT ATTGTTAATG CGGATGTTT 4980
 AGGGCTATCA GTTCCGGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG 5040
 TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTAT 5100
 TACTGGTCGT GTGACTGGTG AATCTGCCAA TGTAAATAAT CCATTTGAGA CGATTGAGCC 5160
 TCAAAATGTA CGTATTTCGA TGAGCGTTTT TCCTGTTGCCA ATGGCTGGCC GTAATATTGT 5220
 TCTGGATATT ACCAGCAAGG CGGATAGTTT GAGTTCTTCT ACTCAGGCCA GTGATGTTAT 5280

TACTAATCAA AGAAGTATTG CTACAACGGT TAATTTGGGT GATGGACAGA CTCTTTACT	5340
CCGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGGTACCGT TCCCTGTCTAA	5400
AATCCCTTTA ATCGGCCTCC TGTTAGCTC CCGCTCTGAT TCCAACGAGG AAACCAACGTT	5460
ATACGTGCTC GTCAAAGCAA CCATACTACG CGCCCTGTAG CGGGCGATTA AGCCGGCGG	5520
GTGTGGTGGT TACCGGCACC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCCT	5580
TCGCTTTCTT CCCTTCCTT CTGGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC	5640
GGGGGCTCCCC TTAGGGTTC CGATTAGTG CTTTACGGCA CCTCGACCCC AAAAAACTTG	5700
ATTTGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA	5760
CGTTGGAGTC CACGTTCTT AATAGTGGAC TCTTGTCCA AACTGGAACA ACACCTCAACC	5820
CTATCTCGGG CTATTCTTT GATTATAAG GGATTTGCC GATTTCGGAA CCACCATCAA	5880
ACAGGATTTC CGCCTGCTGG GGCAAACCGAG CGTGGACCCCG TTGCTGCAAC TCTCTCAGGG	5940
CCAGGCGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTCAAAAGAA AAACCACCC	6000
GGCCGCCAAT ACCCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TGCAGCTGGC	6060
ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC	6120
TCACTCATTA GGCAACCCAG GCTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA	6180
TTGTGAGCGG ATAACAATTT CACACCGTC ACTTGGCACT GGCCGTCGTT TTACAACGTC	6240
GTGACTGGGA AAACCCCTGGC GTTACCCAAG CTTTGTACAT GGAGAAAATA AAGTGAACAA	6300
AAGCACTATT GCACCTGGCAC TCTTACCGTT ACTGTTTACCC CCTGTGGCAA AAGCCCAGGT	6360
CCAGCTGCTC GAGTCGGTCT TCCCCCTGGC ACCCTCCTCC AAGAGCACCT CTGGGGCAC	6420
AGCGGCCCTG GGCTGCCCTGG TCAAGACTAA TTCCCCGAAC CGGTGACGGT GTCGTGGAAC	6480
TCAGGGCCCC TGACCAAGCGG CGTGCACACC TTCCCCGGCTG TCCTACAGTC CTCAGGACTC	6540
TACTCCCTCA GCAGCGTGGT GACCGTGCC TCCAGCAGCT TGGGCACCCA GACCTACATC	6600
TGCAACGTGA ATCACAAGCC CAGCAACACC AAGGTGGACA AGAAAGCAGA GCCCAAATCT	6660
TGTACTAGTG GATCCTACCC GTACGACGTT CCGGACTACCG CTTCTTAGGC TGAAGGGCAT	6720
GACCCTGCTA AGGCTGCATT CAATAGTTTA CAGGCAAGTG CTACTGAGTA CATTGGCTAC	6780
GCTTGGGCTA TGGTAGTAGT TATACTTGGT GCTACCATAG GGATTAATT ATTCAAAAG	6840
TTTACGAGCA AGGCTTCTTA AGCAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC	6900
AGTTGGCAG CCTGAATGGC GAATGGCGCT TTGCTGGTT TCCGGCACCA GAAGGGTGC	6960
CGGAAAGCTG GCTGGAGTGC GATCTTCCCTG AGGCCGATAC GGTGGTCGTC CCCTCAAAC	7020
GGCAGATGCA CGGTTACGAT CGGCCCATCT ACACCAACGT AACCTATCCC ATTACGGTCA	7080
ATCCGCCGTT TGTCCCACG GAGAATCCGA CGGGTTGTTA CTCGCTCACA TTTAATGTTG	7140
ATGAAAGCTG CCTACAGGAA GCCCAGACCC GAATTATTT TGATGGCCTT CCTATTGGTT	7200
AAAAAAATGAG CTGATTTAAC AAAAATTAA CGCGAATTTC AACAAAATAT TAACGTTTAC	7260
AATTTAAATA TTGCTTATA CAATCTTCCCT GTTTTGGGG CTTTCTGAT TATCAACCGG	7320

GGTACATATG ATTGACATGC TAGTTTACG ATTACCCCTTC ATCGATTCTC TTGTTTGCTC 7380
 CAGACTCTCA GGCATGACC TGATAGCCTT TGTAGATCTC TCAAAAATAG CTACCCCTCTC 7440
 CGGCATTAAT TTATCAGCTA GAACGGTTCA ATATCATATT GATGGTGATT TGACTGTCTC 7500
 CGGCCTTCT CACCCTTTTC AATCTTAC TACAGATTAC TCAGGCATTG CATTAAAAT 7560
 ATATGAGGGT TCTAAAAATT TTATCCTTG CGTTGAAATA AAGGCTCTC CGGCAAAAGT 7620
 ATTACAGGGT CATAATGTTT TTGGTACAAC CGATTTAGCT TTATGCTCTG AGGCTTTATT 7680
 GCTTAATTT GCTAATTCTT TGCCTTGCCCT GTATGATTAA TTGGACGTT 7729

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7557 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGGCCCC AAATGAAAAT 60
 ATAGCTAAAC AGGTATTGAA CCATTTGCCA AATGTATCTA ATGGTCAAAC TAAATCTACT 120
 CGTTCGCAGA ATTGGGAATC AACTGTACA TGGAAATGAAA CTTCCAGACA CCGTACTTTA 180
 GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTG AGCAATTAAAG CTCTAAGCCA 240
 TCCGCAAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCCTGACCTG 300
 TTGGAGTTTG CTTCCGGTCT GGTTGGCTTT GAAGCTCGAA TTAAAACCGG ATATTTGAAG 360
 TCTTTCGGGC TTCCCTTTAA TCTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT 420
 CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACG GTTAAAGCA 480
 TTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGAG TATTGGACGC TATCCAGTCT 540
 AAACATTTA CTATTACCCC CTCTGGAAA ACTTCTTTTG CAAAAGCCTC TCGCTATTCTT 600
 GGTTTTTATC GTCGTCTGGT AAACCGAGGGT TATGATAGTG TTGCTCTTAC TATCCCTCGT 660
 AATTCCTTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCTAA ATCTCAACTG 720
 ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT 780
 TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCCATA AGGTAATTCA 840
 CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTC 900
 CTCGTCAAGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG 960
 AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCCTGGTC 1020
 TGTACACCGT TCATCTGTCC TCTTCAAAAG TTGGTCAGTT CGGTCCCTT ATGATTGACC 1080
 GTCTCCGGCT CGTTCCGGCT AAGTAACATG GAGCAGGTG CGGATTTGCA CACAATTAT 1140
 CAGGGCATGA TACAAATCTC CGTTGTACTT TCTTCCGGC TTGGTATAAT CGCTGGGGGT 1200
 CAAAGATGAG TCTTTAGTG TATTCTTCTG CCTCTTTCTG TTTAGCTTGG TGCCTTCTGTA 1261

GTGGCATTAC	GTATTTACC	CGTTAACCG	AAACTTCCTC	ATGAAAAAGT	CTTAGTCCT	1320
CAAAGCCTCT	GTAGCCGTTG	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA	1380
CGATCCCGCA	AAAGCCGCCT	TTAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	1440
TGCGTCGGCG	ATGGTTGTTG	TCATTGTCGG	CCCAACTATC	GGTATCAAGC	TGTTTAAGAA	1500
ATTCACTCTG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCCTTT	1560
TTTTTGGAGA	TTTCAACGT	GAAAAAATTA	TTATTCGCAA	TTCCCTTAGT	TGTTCCCTTC	1620
TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTAGCAA	AACCCCATAC	AGAAAATTCA	1680
TTTACTAACG	TCTGGAAAGA	CGACAAAAC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
CTGTGGAATG	CTACAGGCGT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA	1800
TGGGTTCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860
TCTGAGGGTG	GGGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGACTACGG	TGATACACCT	1920
ATTCCGGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACTT	ATCCGGCTGG	TACTGAGCAA	1980
AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGG	CTCTTAATAC	TTTCATGTTT	2040
CAGAATAATA	GGTTCCGAAA	TAGGCAGGGG	GCATTAACTG	TTTATACGGG	CACTGTTACT	2100
CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG	2160
TATGACGCTT	ACTGGAACGG	TAAATTAGA	GAATGGCCTT	TCCATTCTGG	CTTTAATGAA	2220
GATCCATTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCGCTAACCC	TCCTGTCAAT	2280
GCTGGCGGCG	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340
GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	TGGTCCGGT	2400
GATTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCAT	2460
GAAAACGGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTGGCTAC	TGATTACGGT	2520
GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT	2580
GGTGATTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCCG	GTGACGGTGA	TAATTACACCT	2640
TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT	2700
TTTGTCTTIA	GGCGCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTGACAA	AATAAAACTTA	2760
TTCCCGTGGTG	TCTTTCGCGTT	TCTTTATAT	GTGCCACCT	TTATGTATGT	ATTTCTACG	2820
TTTGCTAAC	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCGT	2880
TATTATTGCG	TTTCCCTCGGT	TTCCCTCTGG	TAACCTTGTT	CGGCTATCTG	CTTACTTTTC	2940
TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CCTGTTCTT	GCTCTTATTA	TTGGGCTTAA	3000
CTCAATTCTT	GTGGGTTATC	TCTCTGATAT	TAGCGCTCAA	TTACCCCTCG	ACTTTGTCA	3060
GGGTGTTCA	TTAATTCTCC	CGTCTAATGC	GCTTCCCTGT	TTTATGTAA	TTCTCTCTGT	3120
AAAGGCTGCT	ATTTCTACG	TTGACGTTAA	ACAAAAAAATC	GTTCCTTATT	TGGATTGGGA	3180
TAAATAATAT	GGCTCTTAT	TTTGTAACTG	GCAAATTAGG	CTCTGGAAAG	ACGCTCGTTA	3240
GGCTTGGTAA	GATTCAAGGAT	AAAATTGTAG	CTGGGTGCAA	AATAGCAACT	AAATCTTGATT	3300

TAAGGCTTCA	AAACCTCCCG	CAAGTCGGGA	GGTTCGCTAA	AACGCCCTCGC	GTTCTTAGAA	3360
TACCCGATAA	GCCTTCTATA	TCTGATTGCG	TTGCTATTGG	CCCCGGTAAT	GATTCCCTACG	3420
ATGAAAATAA	AAACGGCTTG	CTTGGTCTCG	ATGAGTGCAG	TACTTGGTTT	AATACCCGTT	3480
CTTGGAAATGA	TAAGGAAAGA	CAGCCGATTA	TTGATTGGTT	TCTACATGCT	CGTAAATTAG	3540
CATGGGATAT	TATTTTCTT	GTTCAGGACT	TATCTATTGT	TGATAAACAG	GCGCGTCTG	3600
CATTAGCTGA	ACATGTTGTT	TATTGTCGTC	CTCTGGACAG	AATTACTTTA	CCTTTGTGCG	3660
GTACTTTATA	TTCTCTTATT	ACTGGCTCGA	AAATGCCCT	GCCTAAATTAA	CATGTTGGCG	3720
TTGTTAAATA	TGGCGATTCT	CAATTAAGCC	CTACTGTTGA	GGCTTGGCTT	TATACTGGTA	3780
AGAATTGGTA	TAACGCATAT	GATACTAAAC	AGGCTTTTC	TAGTAATTAT	GATTCCGGTG	3840
TTTATTCTTA	TTTAACGCCCT	TATTTATCAC	ACGGTCGGTA	TTTCAAACCA	TTAAATTAG	3900
GTCAGAAGAT	GAAGCTTA	ACTAAATATT	TGAAAAAGTT	TTCACCGCTT	CTTTGTCTG	3960
CGATTGGATT	TGCATCAGCA	TTTACATATA	GTATATAAC	CCAACCTAAG	CCGGAGGTTA	4020
AAAAGGTAGT	CTCTCAGACC	TATGATTTG	ATAAATTGAC	TATTGACTCT	TCTCAGCGTC	4080
TTAATCTAAG	CTATCGCTAT	GTGTTCAAGG	ATTCTAAGGG	AAAATTAAATT	AATAGCCAGC	4140
ATTTACAGAA	GCAAGGTTAT	TCACTCACAT	ATATTGATT	ATGTA	CTGTT	4200
AAGGTAAATC	AAATGAAATT	GTTAAATGTA	ATTAATTGTT	TTTCTTGTAT	GTGTTCA	4260
TCATCTTCTT	TTGCTCAGGT	AATTGAAATG	AATAATTCCG	CTCTGCCGGA	TTTGTAACT	4320
TGGTATTCAA	ACCAATCAGG	CGAATCCGTT	ATTGTTCTC	CCGATGTAAA	AGGTACTGTT	4380
ACTGTATATT	CATCTGACGT	TAACACGAA	AATCTACGCA	ATTCCTTAT	TTCTGTTTAA	4440
CGTGCTAATA	ATTTGATAT	GGTGGTTCA	ATTCTTCCA	TAATTGAGAA	GTATAATCCA	4500
AACAATCAGG	ATTATATTGA	TGAATTGCCA	TCATCTGATA	ATCAGGAATA	TGATGATAAT	4560
TCCGCTCCTT	CTGGTGGTTT	CTTGTCCG	CAAATGATA	ATGTTACTCA	AACTTTAAA	4620
ATTAATAACG	TTGGGGCAA	GGATTTAATA	CGACTTGTG	AATTGTTGT	AAAGTCTAAT	4680
ACTTCTAAAT	CCTCAAATGT	ATTATCTATT	GACGGCTCTA	ATCTATTAGT	TGTTAGTGCA	4740
CCTAAAGATA	TTTGTAAATA	CTTCCCTCAA	TTCCCTTCTA	CTGTTGATT	CCCAACTGAC	4800
CAGATATTGA	TTGAGGGTTT	GATATTGAG	GTTCAGCAAG	GTGATGCTTT	AGATTTTCA	4860
TTTGCTGCTG	CCTCTCAGCG	TGGCACTGTT	CCAGGGCGGTG	TTAATACTGA	CCGCCTCACC	4920
TCTGTTTAT	CTTCTGCTGG	TGGTTCGTT	GGTATTTTTA	ATGGCGATGT	TTTAGGGCTA	4980
TCAGTTCCGG	CATTAAGAC	TAATAGCCAT	TCAAAATAT	TGTCTGTGCC	ACGTATTCTT	5040
ACGCTTTCAG	GTCAGAAGGG	TTCTATCTCT	GTGCCCAGA	ATGTCCTTT	TATTACTGGT	5100
CGTGTGACTG	GTGAATCTGC	CAATGAAAT	AATCCATTTC	AGACGATTGA	CCGTCAAAAT	5160
GTAGGTATT	CCATGAGCGT	TTTCTGTT	CCAATGGCTG	GGGTAATAT	TGTTCTGGAT	5220
ATTACCAAGCA	AGGCCGATAG	TTTGAGTTCT	TCTACTCAGG	CAAGTGATGT	TATTACTAAT	5280
CAAAGAAGTA	TTGCTACAAAC	GGTTAATTG	CCTGATGGAC	AGACTCTT	ACTCGGTGGC	5340

CTCACTGATT ATAAAAACAC TTCTCAAGAT TCTGGCGTAC CGTTCCTGTC TAAAATCCCT	5400
TTAATCGGCC TCCTGTTAG CTCCCCCTCT GATTCCAACG AGGAAAGCAC GTTATACGTG	5460
CTCGTCAAAG CAACCATAGT ACGCCCCCTG TAGCGGGCGA TTAAGCGCGG CGGGTGTGCT	5520
GGTTACGCCG AGCGTGACCG CTACACTTGC CAGCGCCCTA CGCCCGCTC CTTTCCGCTT	5580
CTTCCCTTCC TTCTCGCCA CGTTCGGGGG CTTTCCCCGT CAAGCTCTAA ATCGGGGGCT	5640
CCCTTACGGG TTCCGATTAA GTGCTTACG GCACCTCGAC CCCAAAAAAC TTGATTTGGG	5700
TGATGGTCA CGTAGTGGGC CATCGCCCTG ATAGACGGTT TTTCGCCCC TTGACGTTGGA	5760
GTCCACGGTC TTTAATAGTG GACTCTTGTG CCAAACGTGA ACAACACTCA ACCCTATCTC	5820
GGGCTATTCT TTGATTTAT AAGGGATTTT GCCGATTTCG CAACCACCGAT CAAACAGGAT	5880
TTTCGCTTGC TGGGGCAAAC CAGCGTGGAC CGCTTGTGTC AACTCTCTCA GGGCCAGGCG	5940
GTGAAGGGCA ATCAGCTGTT GCCCGTCTCG CTGGTAAAAA GAAAAACAC CCTGGCCCC	6000
AATACGCAAAC CGCCCTCTCC CGCGCCGTTG GCCGATTCAAT TAATGCAGCT GGCACGACAG	6060
GTTCGGAC TGGAAAGCGG CGAGTGAGCG CAACGCAATT AATGTGAGTT AGCTCACTCA	6120
TTAGGCACCC CAGGCTTAC ACTTTATGCT TCCGGCTCGT ATGTTGTG TGAAATTGTGAG	6180
CGGATAACAA TTTCACACCG CAAGGAGACA GTCATAATGA AATACTATT CCCTACGGCA	6240
GCCGCTGGAT TGTATTACT CGCTGCCAA CCAGCCATGG CCCAGCTCTT CCCGCCATCT	6300
GATGAGCAGT TGAAATCTGG AACTGCCCTCT GTTGTGTGCC TGCTGAATAA CTTCTATCCC	6360
AGAGAGGCCA AAGTACAGTG GAAGGTGGAT AACGCCCTCC AATCGGGTAA CTCCCAGGAG	6420
AGTGTACACAG AGCAGGACAG CAAGGACAGC ACCTACAGCC TCAGCAGCAC CCTGACGGCTG	6480
AGCAAAGCAG ACTACCGAGAA ACACAAAGTC TACGCCCTGG AAGTCACCCA TCAGGGCCTG	6540
AGCTCGCCCC TCACAAAGAG CTTCAACAGG GGAGAGTGTGTT CTAGAACGGG TCACTTGGCA	6600
CTGGCCGTGCG TTTTACAACG TCGTACTGG GAAAACCTG CGCTTACCCA AGCTTAATCG	6660
CCTTGCAGAA TTCCCTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC	6720
TTCCCAACAG TTGGCCAGCC TGAATGGCGA ATGGCGCTTT GCCTGGTTTC CGGCACCAAGA	6780
AGCGGTGCCG GAAAGCTGGC TGGAGTGGCA TCTTCCTGAG GCCGATAACGG TCGTCGTCCC	6840
CTCAAACCTGG CACATGGACCG CTTACGATGC GCCCATCTAC ACCAACGTAAC CCTATCCCAT	6900
TACGGTCAAT CCGCCGTTTG TTCCCACGGA GAATCCGACG GGTTGTACT CGCTCACATT	6960
TAATGTTGAT GAAAGCTGGC TACAGGAAGG CCAGACGGCA ATTATTTTG ATGGCGTTCC	7020
TATTGGTTAA AAAATGAGCT GATTTAACAA AAATTTAACCG CGAATTTAA CAAAATATTAA	7080
ACGTTTACAA TTAAATATT TGCTTATACA ATCTTCCGT TTTTGGGGCT TTCTGATTA	7140
TCAACCGGGG TACATATGAT TGACATGCTA GTTTTACGAT TACCGTTCAAT CGATTCTCTT	7200
GTTCGCTCCA GACTCTCAGG CAATGACCTG ATAGCCTTTC TAGATCTCTC AAAAATAGCT	7260
ACCCCTCTCCG GCATTAATTATCAGCTAGA ACGGTTGAAT ATCATATTGA TGTCGATTG	7320
ACTGTCTCCG GCCTTCTCA CCCTTTGAA TCTTACCTA CACATTACTC AGGCATTGCA	7381

TTTAAAATAT ATGAGGGTTC TAAAAAATTT TATCCTTGGG TTGAAATAAA GGCTTCTCCC 7440
 CCAAAAGTAT TACAGGGTCA TAATGTTTTT GGTACAACCG ATTTAGCTTT ATGCTCTGAG 7500
 GCTTTATTGC TTAATTTGC TAATTCTTTC CCTTGCCTGT ATGATTATT GGATGTT 7557

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8118 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGGCCCC AAATGAAAAT 60
 ATAGCTAAAC AGGTATTGA CCATTGCGA AATGTATCTA ATGGTAAAC TAAATCTACT 120
 CGTTGCAGA ATTGGGAATC AACTGTTACA TCGAATGAAA CTTCCAGACA CCGTACTTTA 180
 GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATT AGCAATTAAG CTCTAAGCCA 240
 TCTGAAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG 300
 TTGGAGTTG CTTCCGGTCT GGTCGCTTT GAAGCTCGAA TTAAAACGGG ATATTGAAAG 360
 TCTTCGGGC TTCCCTTAA TCTTTTGAT GCAATCCGCT TTGCTCTGA CTATAATAGT 420
 CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACG GTTTAAAGCA 480
 TTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGAG TATTGGACGC TATCCAGTCT 540
 AAACATTTA CTATTACCCC CTCTGGAAA ACTTCTTTG CAAAAGCCTC TCGCTATTTT 600
 GGTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCCTCGT 660
 AATTCCCTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCTAA ATCTCAACTC 720
 ATGAATCTT CTACCTGAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT 780
 TCTTCCCAAC GTCCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA 840
 CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGT 900
 CTCCGTAGGG CAAGCCTTAT TCACTGAATG ACCAGCTTTG TTACGTTGAT TTGGGTAATG 960
 AATATCCGGT TCTTGTCAAG ATTACTCTTC ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC 1020
 TGTACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC 1080
 GTCTGCCCT CGTTCCGGCT AAGTAACATG GACCAAGTCC CGGATTTGGA CACAATTAT 1140
 CAGGGGATGA TACAAATCTC CGTTGTACTT TCTTCCGGC TTGGTATAAT CGCTGGGGGT 1200
 CAAAGATGAG TCTTTAGTG TATTCTTTCG CCTCTTTCTG TTTAGGTTGG TGCCTTCGTA 1260
 GTGGCATTAC GTATTTTACCG GTTTAATGG AAACCTTCCTC ATGAAAAAGT CTTAGTCCT 1320
 CAAAGCCTCT GTAGCCGTTG CTACCCCTCGT TCCGATGCTG TCTTCCGCTG CTGAGGGTGA 1380
 CGATCCCCCA AAAGCCGCCT TTAACCTCCCT GCAAGCCTCA GCGACCCGAAT ATATCGGTTA 1440
 TGCCTGGGGC ATGGTTGTTG TCATTCTCGG CGCAACTATC GGTATCAAGC TCTTTAAAGAA 1500

ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTT	GGAGCCTTT	1560	
TTTTGGAGA	TTTCAACGT	GAAAAAATT	TTATTCGAA	TTCCCTTAGT	TGTTCCCTTC	1620	
TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA	1680	
TTTACTAACG	TCTGGAAAGA	CCACAAA	ACT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
CTGTGGAATG	CTACAGGC	GT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTC	TTACGGTACA	1800
TGGGTTCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860	
TCTGAGGGTG	GGGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT	1920	
ATTCCGGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACTT	ATCCGGCTGG	TACTGAGCAA	1980	
AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT	2040	
CAGAATAATA	GGTCCGAAA	TAGGCAGGGG	GCATTAACTG	TTTATACGGG	CACTGTTACT	2100	
CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG	2160	
TATGACGCC	TTTGTGAATA	TCAAGGCCAA	TCGGTCTGACC	TGCCCTCAACC	TCCTGTCAAT	2220	
GATCCATT	GGCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2280	
GCGGTTCTG	AGGGTGGCGG	CTCTGAGGCA	GGCGGTTCCG	GTGGTGGCTC	TCGTTCCGGT	2340	
GATTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	2400	
GAAAACCCGC	TACAGTCTGA	CCCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	2460	
GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT	2520	
GGTGATTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCAACCT	2580	
TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTGA	ATGTCGCCCT	2640	
TTTGTCTTIA	GGCCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTACAA	AATAAACCTTA	2700	
TTCCGTGGTG	TCTTTCGTT	TCTTTTATAT	GTGCCACCT	TTATGTATGT	ATTTCTACCG	2760	
TTTGCTAAC	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCGT	2820	
TATTATTGCG	TTTCCTCGGT	TTCTTCTGG	TAACTTTGT	CGGCTATCTG	CTTACTTTC	2880	
TTAAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTCATT	GTTCCTTGCT	CTTATTATTG	2940	
GGCTTAACTC	AATTCTTGTG	CGTTATCTCT	CTGATATTAG	CGCTCAATT	CCCTCTGACT	3000	
TTGTTCA	GGGGTCA	ATTCTCCCGT	CTAATGCCCT	TCCCTGTTT	TATGTTATT	3060	
TCTCTGTAAA	GGCTGCTATT	TTCA	TTTGTGTTT	ACGTTAAACA	AAAAATCGTT	3120	
ATTGGGATAAA	ATAATATGGC	TGTTTATTTT	CTAACTGGCA	AATTAGGCTC	TGGAAAGACG	3180	
CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGC	AAAATACCAACTAAT	3240	
CTTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	GTGGGAGGT	TCGCTAAAAC	CCCTCGCGTT	3300	
CTTACAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGCGG	CGCTAATGAT	3360	
TCCTACCGATG	AAAATAAAA	CGGCTTGCTT	GTTCTCGATG	AGTGC	GGGTAC	3420	
ACCCGTTCTT	CGAATGATAA	CGAAAGACAG	CCCATTATTG	ATTGGTTCT	ACATGCTCCT	3480	
					3540		

AAATTAGGAT CGGATATTAT TTTCTTGT CAGGACTTAT CTATTGTTCA TAAACAGGGC 3600
 CGTTCTGCAT TAGCTGAACA TGTTCCTTAT TGTCTCGTC TGGACAGAAT TACTTACCT 3660
 TTTGTCGGTA CTTTATATTTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT 3720
 GTTGGCCTTG TAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTAT 3780
 ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTCTAG TAATTATGAT 3840
 TCCGGTGT TT ATTCTTATT ACGCCTTAT TTATCACAGG GTCGGTATT CAAACCATT 3900
 AATTTAGGTC AGAACATGAA CCTTACTAAA ATATAATTCA AAAAGTTTC ACGCGTTCTT 3960
 TGTCTTGCCTA TTGGATTTCG ATCAGCATT ACATATAGTT ATATAACCCA ACCTAACCG 4020
 GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGACTCTTCT 4080
 CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAA 4140
 AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTIATG TACTGTTCC 4200
 ATTAAAAAAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTGTTT TCTTGATGTT 4260
 TGTTTCATCA TCTTCTTGT CTCAGGTAAT TGAAATGAAT AATTCCCTC TGCGCGATT 4320
 TGTAACCTGG TATTCAAAGC AATCAGGCGA ATCCGTTATT GTTCTCCCG ATGTAAGG 4380
 TACTGTTACT GTATATTCACT GACGTTAA ACCTGAAAAT CTACGCAATT TCTTATTTCT 4440
 TGTTTACGT GCTAATAATT TTGATATGGT TGTTCAATT CCTTCATAA TTCAAGAGTA 4500
 TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA 4560
 TGATAATTCC GCTCCTTCTG GTGGTTCTT TGTTCCGAA AATGATAATG TTACTCAAAC 4620
 TTTTAAATT AATAACGTTG GGGCAAAGGA TTTAATACGA GTTGTGAAAT TGTTGTA 4680
 GTCTAAACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTACTGTT 4740
 TAGTGCACCT AAAGATATT TAGATAACCT TCCTCAATTCTTCTACTG TTGATTGCC 4800
 AACTGACCAAG ATATTGATTG AGGGTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTAA 4860
 TTTTCAATT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG 4920
 CCTCACCTCT GTTTATCTT CTGCTGGTGG TTGTTGCGT ATTGTTAATG GCGATGTTT 4980
 AGGGCTATCA GTTCCGCGAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG 5040
 TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTAT 5100
 TACTGGTCGT GTGACTGGTG AATCTGCCAA TGAAATAAT CCATTTCAGA CGATTGAGCG 5160
 TCAAAATGTA GGTATTTCGA TGAGCGTTT TCCGTGCA ATGGCTGGCG GTAATATTGT 5220
 TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCAA GTGATGTTAT 5280
 TACTAATCAA AGAAGTATTG CTACAACGGT TAATTGCGT GATGGACAGA CTCTTTACT 5340
 CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCGTCTAA 5400
 AATCCCTTAA ATCGGCCTCC TGTGTTAGCTC CCCCTCTGAT TCCAACGAGG AAAGCACGTT 5460
 ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCCATTAGC AGCGCGGGCG 5520
 GTGTGGTGGT TACGGCGAGC GTGACCGCTA CACTTCCAGS CCCCCCTAGCG CCCCCCTCCTT 5580

TCGCTTCTT CCCTTCCCTT CTGGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC 5640
 GGGGGCTCCC TTTAGGGTTC CGATTAGTG CTTTACGGCA CCTCGACCCC AAAAAACTTG 5700
 ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTGA 5760
 CGTTGGACTC CACGGTCTTT AATAGTGGAC TCTTGTTC AACTGGAACA ACACCTCAACC 5820
 CTATCTCGGG CTATTCTTT GATTATAAG GGATTTGCC GATTCGGAA CCACCATCAA 5880
 ACAGGATTT CGCCTGCTGG CGCAAACCGAG CGTGGACCCGC TTGCTGCAAC TCTCTCAGGG 5940
 CCAGGCGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCC 6000
 GGCCTCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TGCAGCTGGC 6060
 ACCACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCCAA CGCAATTAAAT GTGAGTTAGC 6120
 TCACTCAATTAA GGCACCCAG GCTTACACT TTATGTTCC GGCTCGTATG TTGTGTGGAA 6180
 TTGTGAGCGG ATAACAATTI CACACGCCAA GGAGACAGTC ATAATGAAAT ACCTATTGCC 6240
 TACGGCAGCC GCTGGATTGT TATTACTCGC TGCCCAACCA GCCATGGCCG AGCTCTTCCC 6300
 GCCATCTGAT GAGCAGTTGA AATCTGGAAC TGCCTCTGTT GTGTGCCTGC TGAATAACTT 6360
 CTATCCCAGA GAGGCCAAAG TACAGTGGAA GGTGGATAAC CCCCTCCAAT CGGCTAACTC 6420
 CCAGGAGACT GTCACAGAGC AGGACACCAA GGACAGCACC TACAGCCTCA GCAGCACCC 6480
 GACGCTGAGC AAAGCAGACT ACGAGAAACA CAAAGTCTAC GCCTGCGAAG TCACCCATCA 6540
 GGGCCTGAGC TCGCCCGTCA CAAAGAGCTT CAACAGGGGA GAGTGTCTA GAACGGTCA 6600
 CTTGGCACTG GCCGTGTTT TACAACGTGG TGACTGGAA AACCCCTGGCG TTACCCAAGC 6660
 TTTGTACATG GAGAAAATAA AGTGAACAA ACCACTATTG CACTGGCACT CTTACCGTTA 6720
 CTGTTTACCC CTGTGGAAA AGCCGCCTCC ACCAAGGGCC CATCGTCTT CCCCCCTGGCA 6780
 CCCTCCTCCA AGAGCACCTC TGGGGCACA GCGGCCCTGG GCTGCCCTGGT CAAGACTAAT 6840
 TCCCCGAACC GGTGACGGTG TCGTGGAACT CAGGCGCCCT GACCAGCGGC GTGCACACCT 6900
 TCCCCGGCTGT CCTACAGTCC TCAGGACTCT ACTCCCTCAG CAGCGTGGTG ACCGTGCCCT 6960
 CCACCAAGCTT GGGCACCCAG ACCTACACT GCAACGTGAA TCACAAGCCC AGCAACACCA 7020
 AGGTGGACAA GAAAGCAGAG CCCAAATCTT GTACTAGTGG ATCCTACCCG TACGACGTTC 7080
 CGGACTACGGC TTCTTACGGCT GAAGGGCATG ACCCTGCTAA GGCTGCATTTC AATAGTTTAC 7140
 AGGCAAGTGC TACTGAGTAC ATTGGCTACG CTTGGCTAT GGTAGTAGTT ATAGTTGGTG 7200
 CTACCATAGG GATTAAATTAA TTCAAAAAGT TTACCGAGCAA GGCTTCTTAA CCAATAGCGA 7260
 AGAGGCCCGC ACCGATCGCC CTTCCCAACCA GTTGCAGC CTGAATGGCG AATGGCGCTT 7320
 TGCCTGGTTT CCGGCACCCAG AAGGGTCCC CGAAAGCTGG CTGGAGTGGC ATCTTCTGA 7380
 CGCCGATACG GTCGTCGTCC CCTCAAACCTG CGAGATGCAC GGTTACGATG CGCCCATCTA 7440
 CACCAACGTA ACCTATCCCA TTACGGTCAA TCCCCGGTTT GTTCCCACGG AGAATCCGAC 7500
 CGGTGTTAC TCGCTCACAT TTAATGTTGA TGAAAGCTGG CTACAGGAAG GCCAGACCCG 7560
 AATTATTTT GATGGCGTTC CTATTGGTTA AAAATGAGC TGATTTAAC AAAATTTAAC 7620

CGCAATTATA ACAAAATATT AACGTTTACA ATTAAATAT TTGCTTATAC AATCTTCCTG 7680
 TTTTGGGC TTTCTGATT ATCAACCGGG GTACATATGA TTGACATGCT AGTTTACGA 7740
 TTACCGTTCA TCGATTCTCT TGTTGCTCC AGACTCTAG GCAATGACCT GATAGCCCTT 7800
 GTAGATCTCT CAAAAATAGC TACCTCTCC GGCATTAATT TATCAGCTAG AACGGTTGAA 7860
 TATCATATTG ATGGTGAATT GACTGCTCC GGCCTTCTC ACCCTTTGA ATCTTACCT 7920
 ACACATTACT CAGGCATTGC ATTAAAATA TATCAGGGTT CTAAAAATT TTATCCTG 7980
 GTTGAATAA AGGCTCTCC CGCAAAAGTA TTACAGGGTC ATAATGTTT TGGTACAACC 8040
 GATTTAGCTT TATGCTCTGA GGCTTATTG CTTAATTG CTAATTCTT GCCTTGCCTG 8100
 TATGATTAT TGGACGTT 8118

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(5, "")
- (D) OTHER INFORMATION: /note- "S REPRESENTS EQUAL MIXTURE OF G AND C"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(6, "")
- (D) OTHER INFORMATION: /note- "M REPRESENTS EQUAL MIXTURE OF A AND C"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(8, "")
- (D) OTHER INFORMATION: /note- "R REPRESENTS EQUAL MIXTURE OF A AND G"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(11, "")
- (D) OTHER INFORMATION: /note- "K REPRESENTS EQUAL MIXTURE OF G AND T"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(20, "")
- (D) OTHER INFORMATION: /note- "W REPRESENTS EQUAL MIXTURE OF A AND T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGGTSMARCT KCTCGAGTCW GG

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGGTCCAGCT GCTCGACTCT GG

(2) INFORMATION FOR SEQ ID NO:8:

22

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AGGTCCAGCT GCTCGACTCA GG

(2) INFORMATION FOR SEQ ID NO:9:

22

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AGGTCCAGCT TCTCGACTCT GG

(2) INFORMATION FOR SEQ ID NO:10:

22

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGGTCCAGCT TCTCGACTCA GG

(2) INFORMATION FOR SEQ ID NO:11:

22

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

64

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AGGTCCAACT GCTCGAGTCT GG

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGGTCCAACT GCTCGAGTCA GG

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGGTCCAACT TCTCGAGTCT GG

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGGTCCAACT TCTCGAGTCA GG

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: misc difference
- (B) LOCATION: replace(5..6, "")
- (D) OTHER INFORMATION: /note- "N-INOSINE"

(ix) FEATURE:

- (A) NAME/KEY: misc difference
- (B) LOCATION: replace(8, "")
- (D) OTHER INFORMATION: /note- "N-INOSINE"

22

22

22

22

65

(ix) FEATURE:

- (A) NAME/KEY: misc_diffr.ence
- (B) LOCATION: replace(11, "")
- (D) OTHER INFORMATION: /note- "N-INOSINE"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(20, "")
- (D) OTHER INFORMATION: /note- "W REPRESENTS EQUAL MIXTURE OF A AND T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AGGTNNANCT NCTCGAGTCW GG

22

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTATTAAC TAACTA GTAACCGTAA CAGTGGTGCC TTGCCCCA

38

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AGGCTTACTA GTACAATCCC TGGGCACAAT

30

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCAGTTCCGA GCTCGTTGTG ACTCAGGAAT CT

32

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCAGTTCCGA GCTCGTGTG ACCGAGCCGC CC

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCAGTTCCGA GCTCGTGCTC ACCCAGTCTC CA

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCAGTTCCGA GCTCCAGATG ACCCAGTCTC CA

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCAGATGTGA GCTCGTGATG ACCCAGACTC CA

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCAGATGTGA GCTCGTCATG ACCCAGTCTC CA

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCAGTTCCGA GCTCGTGATG ACACAGTCTC CA

(2) INFORMATION FOR SEQ ID NO:25:

32

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCAGCATTCT AGAGTTTCAG CTCCAGCTTC CC

(2) INFORMATION FOR SEQ ID NO:26:

32

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GGGCCGTCTA GAATTAACAC TCATTCCTGT TGAA

(2) INFORMATION FOR SEQ ID NO:27:

34

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GATCCTAGGC TGAAGGCGAT GACCCCTGCTA AGGCTGC

(2) INFORMATION FOR SEQ ID NO:28:

37

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ATTCAATACT TTACAGGCAA GTGCTACTGA CTACA

35

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TTGGCTACGC TTGGGCTATG GTAGTAGTTA TAGTT

35

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GGTGCTACCA TAGGGATTAA ATTATTCAAA AAGTT

35

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TACGAGCAAG GCTTCTTA

18

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

AGCTTAAGAA GCCTTGCTCG TAAACTTTTT GAATAATT

39

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AATCCCTATG GTAGCACCAA CTATAACTAC TACCAT

36

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AGCCCCAAGCG TAGCCAATGT ACTCACTAGC ACTTG

35

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCTGTAAACT ATTGAATGCA GCCTTAGCAG GGTC

34

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

ATCGCCTTCA GCCTAG

16

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CATTTTGCA GATGGCTTAG A

21

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

70

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:
 TAGGATTAAC GTCCAATA

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

18

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:
 ATATATTTA GTAAGCTTCA TCTTCT

(2) INFORMATION FOR SEQ ID NO:40:

26

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:
 GACAAAGAAC GCGTGAAAC TTT

(2) INFORMATION FOR SEQ ID NO:41:

23

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 35 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:
 CGGGGCCTCT TCGCTATTGC TTAAGAAGCC TTGCT

(2) INFORMATION FOR SEQ ID NO:42:

35

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 43 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

AAACGACGGC CAGTGCCAAG TGACGCGTGT GAAATTGTTA TCC

43

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGCGAAAGGG AATTCTGCAA GGCGATTAAG CTTGGGTAAC GCC

43

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GGCGTTACCC AAGCTTGTATGGAGAAA ATAAAG

36

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

TGAAACAAAG CACTATTGCA CTGGCACTCT TACCGTTACC GT

42

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TACTGTTTAC CCCTGTGACA AAAGCCGCC AGGTCCAGCT GC

42

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

72

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TCGACTCAGG CCTATTGTGC CCAGGGATTG TACTAGTGGA TCCG

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

TGGCGAAAGG GAATTCCGAT CCACTAGTAC AATCCCTG

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GGCACAAATAG GCCTGACTCG AGCAGCTGGA CCAGGGCGGC TT

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

TTGTCACAGG CGTAAACAGT AACGGTAACG GTAAGTGTGC CA

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CTGCAATACT GCTTTGTTC ACTTTATTCTT CTCCATGTAC AA

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

44

38

42

42

42

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:
TAACCGCTAAG AGTGCCAGTG C

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:
CACCTTCATG AATTGGCAA GGAGACAGTC AT

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:
AATTGGCAA GGAGACAGTC AT

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:
AATGAAATAC CTATTGCCCTA CGGCAGCCGC TGGATTGTT

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:
ATTACTCGCT GCCCAACCAG CCATGGCCGA GCTCGTGAT

21

32

22

39

39

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GACCCAGACT CCAGATATCC AACAGGAATG AGTGTAAAT

39

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

TCTAGAACGC GTC

13

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TTCAGGTTGA AGCTTACGCG TTCTAGAATT AACACTCATT CCTGT

45

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

TGGATATCTG GAGTCTGGGT CATCACCGAGC TCGGCCATG

39

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

75

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GCTGGTTGGG CAGCGAGTAA TAACAATCCA GCGGCTGCC

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GTAGGCAATA GGTATTTCAT TATGACTGTC CTTGGCG

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

TGACTGTCTC CTTGGCGTGT GAAATTGTTA

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

TAACACTCAT TCCGGATGGA ATTCTGGAGT CTGGGT

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GCCAGTGCCA AGTGACGCCGT TCTA

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

39

37

30

36

24

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

ATATATTTA GTAAGCTTCA TCTTCT

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

26

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GACAAAGAAC CGGTGAAAAC TTT

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

23

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

CTGAACCTGT CTGGGACCAC AGTTGATGCT ATAGGATCAG ATCTAGAATT CATTAGAGA
CTGGCCTGGC TTCTGC

60

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

76

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

TCGACCGTTG GTAGGAATAA TGCAATTAAT GGAGTAGCTC TAAATTAGA ATTCACTAC
ACCCAGTGCA TCCAGTAGCT

60

80

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GGTAAACAGT AACGGTAAGA GTGCCAG

27

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

CGCCTTCAGC CTAAGAACCG TAGTCCGGAA CGTCGTACGG GTAGGATCCA CTAG

54

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

CACCGGTTCG GGGATTAGT CTTGACCAGG CAGCCCAGGG C

41

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

ATTCCACACACA TTATACGAGC CGGAAGCATA AAGTCTCAAG CCTGGGGTCC C

51

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

CTGCTCATCA GATGGCGGGA AGAGCTCGGC CATGGCTGGT TG

42

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

78

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

GAACAGAGTC ACCGAGGGGG CGAGCTCGGC CATGGCTGGT TG

42

I Claim:

1. A composition of matter comprising a plurality of cells containing diverse combinations of first and second DNA sequences encoding first and second polypeptides which form heteromeric receptors, one or both 5 of said polypeptides being expressed as fusion proteins on the surface of a cell.
2. The composition of claim 1, wherein said plurality of cells are E. coli.
3. The composition of claim 1, wherein said heteromeric receptors selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.
4. The composition of claim 1, wherein said first and second DNA sequences encode functional portions of heteromeric receptors.
5. The composition of claim 4, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.
6. The composition of claim 1, wherein said cell produces filamentous bacteriophage.
7. The composition of claim 6, wherein said filamentous bacteriophage are selected from the group consisting of M13, fd and fl.
8. The composition of claim 6, wherein at least one of the encoded first or second polypeptides is expressed as a fusion protein with gene VIII.

9. A kit for the preparation of vectors useful for the coexpression of two or more DNA sequences encoding polypeptides which form heteromeric receptors comprising two vectors, a first vector having two pairs of restriction sites symmetrically oriented about a cloning site which can be combined with a second vector, having two pairs of restriction sites symmetrically oriented about a cloning site and in an identical orientation to that of the first vector, wherein one or both vectors contains sequences necessary for expression of polypeptides encoded by DNA sequences inserted in said cloning sites.

10. The kit of claim 9, wherein said first and second vectors are circular.

11. The kit of claim 9, wherein said expression peptides is as fusion proteins on the surface of a cell.

12. The kit of claim 9, wherein said cell produces filamentous bacteriophage.

13. The kit of claim 9, wherein said filamentous bacteriophage is selected from the group consisting of M13, fd and f1.

14. The kit of claim 13, wherein at least one of the DNA sequences is expressed as a fusion protein with gene VIII.

15. The kit of claim 9, wherein said two pairs of restriction sites are Hind III-Mlu I and Hind III-Mlu I.

16. A cloning system for the coexpression of two or more DNA sequences encoding polypeptides which form a heteromeric receptor, comprising a set of first vectors having a diverse population of first DNA sequences and a 5 set of second vectors having a diverse population second DNA sequences, said first and second vectors having two pairs of restriction sites symmetrically oriented about a cloning site for containing said first and second populations of DNA sequences so as to allow only the 10 operational combination of vector sequences containing said first and second DNA sequences.

17. The cloning system of claim 16, wherein said first and second vectors are circular.

18. The cloning system of claim 16, wherein said heteromeric receptors selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.

19. The cloning system of claim 16, wherein said first and second DNA sequences encode functional portions of heteromeric receptors.

20. The cloning system of claim 19, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.

21. The cloning system of claim 16, wherein said coexpression of two or more DNA sequences encoding polypeptides which form a heteromeric receptor is on the surface of cell.

22. The cloning system of claim 16, wherein said cell produces a filamentous bacteriophage.

23. The cloning system of claim 22 wherein said filamentous bacteriophage selected from the group consisting of M13, fd and f1.

24. The cloning system of claim 23, wherein at least one of the DNA sequences is expressed as a fusion protein with the protein product of gene VIII.

25. The cloning system of claim 16, wherein said two pairs of restriction sites are Hind III-Mlu I and Hind III-Mlu I.

26. A plurality of expression vectors containing a plurality of possible first and second DNA sequences encoding polypeptides which form a heteromeric receptor exhibiting binding activity toward a preselected molecule,
5 said DNA sequence encoding heteromeric receptors being operatively linked to genes encoding surface proteins of a cell.

27. The expression vectors of claim 26, wherein said expression vectors are circular.

28. The expression vectors of claim 23, wherein said heteromeric receptors are selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.

29. The expression vectors of claim 26, wherein said first and second DNA sequences encode functional portions of heteromeric receptors.

30. The expression vectors of claim 29, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.

31. The expression vectors of claim 26, wherein said cells produce filamentous bacteriophage.

32. The expression vectors of claim 26, wherein said filamentous bacteriophage are selected from the group consisting of M13, fd and f1.

33. The expression vectors of claim 32, wherein at least one of the encoded first or second polypeptides is expressed as a fusion protein with gene VIII.

34. A method of constructing a diverse population of vectors capable of expressing a diverse population of heteromeric receptors, comprising:

5 (a) operationally linking to a first vector a first population of diverse DNA sequences encoding a diverse population of first polypeptides, said first vector having two pairs of restriction sites symmetrically oriented about a cloning site;

10 (b) operationally linking to a second vector a second population of diverse DNA sequences encoding a diverse population of second polypeptides, said second vector having two pairs of restriction sites symmetrically oriented about a cloning site in an identical orientation to that of the first vector; and

15 (c) combining the vector products of step (a) and (b) under conditions which allow only the operational combination of vector sequences containing said first and second DNA sequences.

35. The method of claim 34, wherein said first and second vectors are circular.

36. The method of claim 34, wherein said heteromeric receptors are selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.

37. The method of claim 34, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.

38. The method of claim 34, wherein said expression of a diverse population of heteromeric receptors is on the surface of a cell.

39. The method of claim 37, wherein said cell produces a bacteriophage.

40. The method of claim 39, wherein said filamentous bacteriophage is selected from the group consisting of M13, fd and fl.

41. The method of claim 34, wherein at least one of said first or second DNA sequences is expressed as a gene VIII fusion protein.

42. The method of claim 34, wherein said two pairs of restriction sites are Hind III-Mlu I and Hind III-Mlu I.

43. The method of claim 34, wherein said combining step further comprises:

5 (C1) restricting said first vector with a restriction enzyme recognizing one of the restriction sites encoded in said two pairs of restriction sites;

10 (C2) restricting said second vector with a different restriction enzyme recognizing the second restriction site encoded in said two pairs of restriction sites;

(C3) digesting the 3' ends of said restricted first and second vectors with an exonuclease; and

15 (C4) annealing said first and second vectors.

44. A method for selecting a heteromeric receptor exhibiting binding activity toward a preselected molecule from a population of diverse heteromeric receptors, comprising:

5

(a) operationally linking to a first vector a first population of diverse DNA sequences encoding a diverse population of first polypeptides, said first vector having two pairs of restriction sites symmetrically oriented about a cloning site;

10

15

20

25

30

(b) operationally linking to a second vector a second population of diverse DNA sequences encoding a diverse population of second polypeptides, said second vector having two pairs of restriction sites symmetrically oriented about a cloning site in an identical orientation to that of the first vector;

(c) combining the vector products of step (a) and (b) under conditions which allow only the operational combination of vector sequences containing said first and second DNA sequences.

(d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of first and second DNA sequences; and

(e) determining the heteromeric receptors which bind to said preselected molecule.

45. The method of claim 44, wherein said first and second vectors are circular.

46. The method of claim 44, wherein said heteromeric receptors are selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.

47. The method of claim 44, wherein said first and second DNA sequences encode functional portions of heteromeric receptors.

48. The method of claim 47, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.

49. The method of claim 44, wherein said expression of a diverse population of heteromeric receptors is on the surface of a cell.

50. The method of claim 49, wherein said cell produces a filamentous bacteriophage.

51. The method of claim 50, wherein said filamentous bacteriophage is selected from the group consisting of M13, fd and f1.

52. The method of claim 51, wherein at least one of said first or second DNA sequences is expressed as a gene VIII fusion protein.

53. The method of claim 44, wherein said two pairs of restriction sites are Hind III-Mlu I and Hind III-Mlu I.

54. The method of claim 44, wherein said combining step further comprises:

5

(C1) restricting said first vector with a restriction enzyme recognizing one of the restriction sites encoded in said two pairs of restriction sites;

10

(C2) restricting said second vector with a different restriction enzyme recognizing the second restriction site encoded in said two pairs of restriction sites;

(C3) digesting the 3' ends of said restricted first and second vectors with an exonuclease; and

15

(C4) annealing said first and second vectors.

5 55. A method for determining the nucleic acid sequences encoding a heteromeric receptor exhibiting binding activity toward a preselected molecule from a diverse population of heteromeric receptors, comprising:

10 (a) operationally linking to a first vector a first population of diverse DNA sequences encoding a diverse population of first polypeptides, said first vector having two pairs of restriction sites symmetrically oriented about a cloning site;

15 (b) operationally linking to a second vector a second population of diverse DNA sequences encoding a diverse population of second polypeptides, said second vector having two pairs of restriction sites symmetrically oriented about a cloning site in an identical orientation to that of the first vector;

20 (c) combining the vector products of step (a) and (b) under conditions which allow only the operational combination of vector sequences containing said first and second DNA sequences.

25 (d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of first and second DNA sequences;

- (e) determining the heteromeric receptors which bind to said preselected molecule;
- 5 (f) isolating the nucleic acid sequences encoding said first and second polypeptides; and
- (g) sequencing said nucleic acid sequences.

56. The method of claim 55, wherein said first and second vectors are circular.

57. The method of claim 55, wherein said first heteromeric receptors selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.

58. The method of claim 55, wherein said first and second DNA sequences encode functional portions of heteromeric receptors.

59. The method of claim 58, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.

60. The method of claim 55, wherein said expression of a diverse population of heteromeric receptors is on the surface of a cell filamentous bacteriophage selected from the group consisting of M13, fd and f1 and at 5 least one of said first or second DNA sequences is expressed as a gene VIII fusion protein.

61. The method of claim 55, wherein said cell produces filamentous bacteriophage.

66. A vector comprising two copies of a gene encoding a filamentous bacteriophage coat protein, one copy of said gene capable of being operationally linked to a DNA sequence encoding a polypeptide of a heteromeric receptor 5 wherein said DNA sequence can be expressed as a fusion protein on the surface of said filamentous bacteriophage or as a soluble polypeptide.

67. The vector of claim 66, wherein said two copies of said gene encode substantially the same amino acid sequence but have different nucleotide sequences.

68. The vector of claim 66, wherein said one copy of said gene is expressed on the surface of said filamentous bacteriophage.

69. The vector of claim 66, wherein said bacteriophage coat protein is M13 gene VIII.

70. The vector of claim 66, wherein said vector has substantially the same sequence as that shown in Figure 2 (SEQ ID NO: 1).

71. A vector comprising sequences necessary for the coexpression of two or more inserted DNA sequences encoding polypeptides which form heteromeric receptors and two copies of a gene encoding a filamentous bacteriophage 5 coat protein, one copy of said gene capable of being operationally linked to one of said two or more inserted DNA sequences wherein said DNA sequence can be expressed as a fusion protein on the surface of said filamentous bacteriophage or as a soluble polypeptide.

72. The vector of claim 71, wherein said two copies of said gene encode substantially the same amino acid sequence but have different nucleotide sequences.

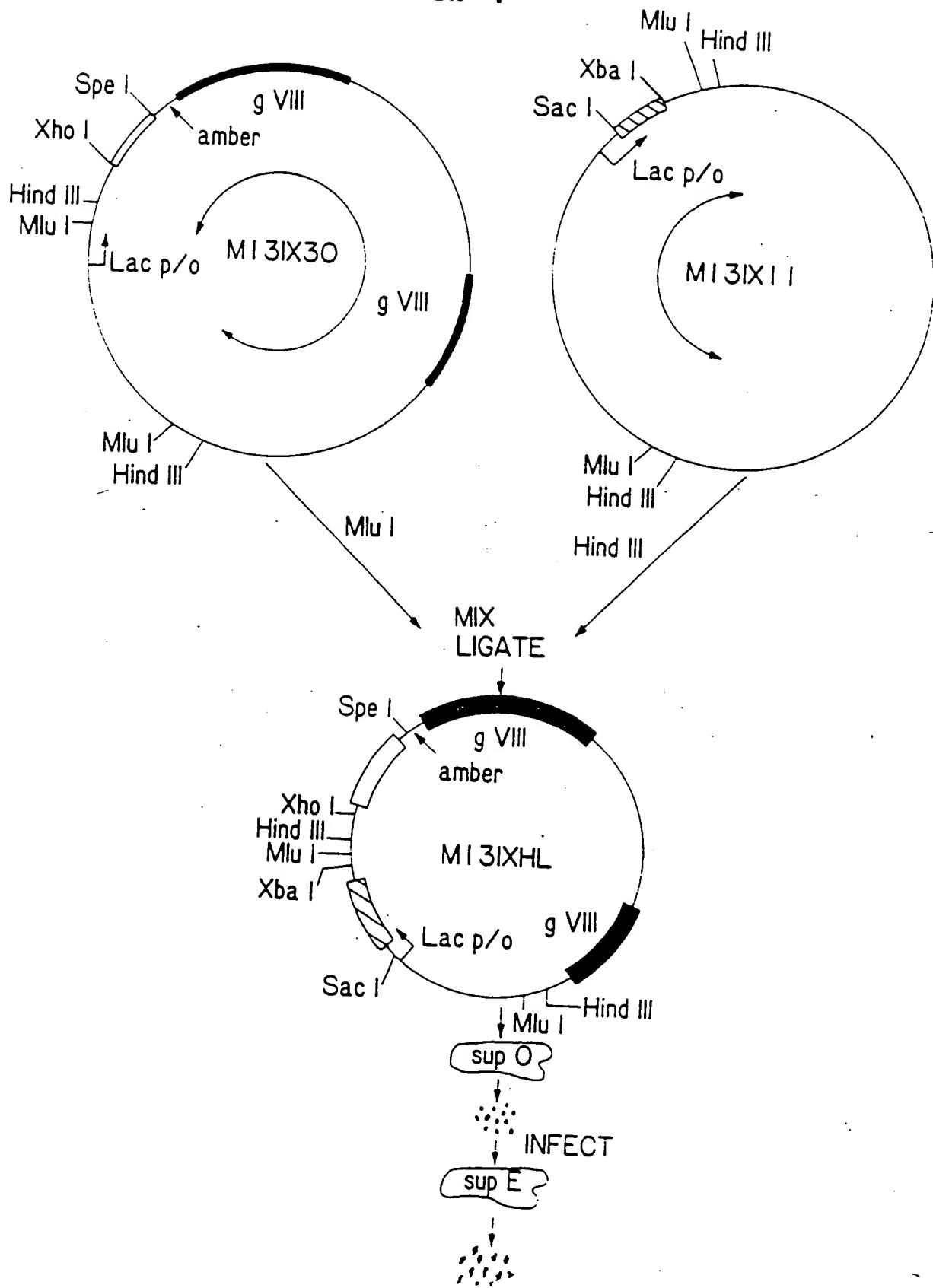
73. The vector of claim 71, wherein said one copy of said gene is expressed on the surface of said filamentous bacteriophage.

74. The vector of claim 71, wherein said bacteriophage coat protein is M13 gene VIII.

75. The vector of claim 71, wherein said vector has substantially the same sequence as that shown in Figure 6 (SEQ ID NO: 5).

1/11

FIG. 1



	1	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCA	CTCGCGCCCC	AAATGAAAAT	60
61	ATAGCTAAAC	AGGTTATTGA	CCATTTGCCA	AATGTATCTA	ATGGTCAAC	AAATCTACT	120
121	CGTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTAACCTTA	180
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAAGATT	AGCAATTAA	CTCTAAGCCA	240
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG	300
301	TTGGAGTTTG	CTTCCGGTCT	GGTTCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTGAAAG	360
361	TCTTCGGGC	TTCCCTTAA	TCTTTTGAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT	420
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGA	GTAAAGCA	480
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT	540
541	AAACATTTA	CTATTACCCC	CTCTGGCAA	ACTTCCTTTG	CAAAGCTCT	TCGCTATTTT	600
601	GGTTTTTATC	GTCGTCTGGT	AAACAGGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCCTCGT	660
661	AATTCCCTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTTG	GTATTCCCAA	ATCTCAACTG	720
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTATTA	CGTAGATTTT	780
781	TCTTCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGATA	AGGTAAATTCA	840
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGT	900
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAAATG	960
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC	1020
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAG	TTGGTCAAGT	CGGGATCCCT	ATGATTGACC	1080
1081	GTCTGCGCCT	CGTTCCGGCT	AAAGTAACATG	GAGCAGGTCG	CGGAGTTTCGA	CACAATTAT	1140
1141	CAGGGCATGA	TACAAATCTC	CGTTGTA	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT	1200
1201	CAAAGATGAG	TGTTTATAGT	TATTCTTCG	CCTCTTCG	TTTAGGTTGG	TGCCCTCGTA	1260
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACTCCCTC	ATGAAAAAGT	CTTAGTCTC	1320
1321	CAAAGCCTCT	GTAGCCGGT	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA	1380
1381	CGATCCCGCA	AAAGCGGCCT	TTAACCTCC	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	1440
1441	TGCGTGGGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTAAGAA	1500
1501	ATTACACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTT	GGAGCCCTTT	1560
1561	TTTTGGGAGA	TTTCAACGT	AAAAAAATTA	TTATTGCAA	TTCTTTAGT	TGTTCCCTTC	1620
1621	TATTCTCACT	CGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	ACCCCATAC	AGAAAATTCA	1680
1681	TTTFACTAACG	TCTGGAAAGA	CGACAAAAC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
1741	CTGTGGAATG	CTACAGGCGT	TGTTAGTTGT	ACTGGTGACG	AAACTCAGT	TTACGGTACA	1800
1801	TGGGTTCTA	TTGGGCTTG	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT	1920
1921	ATTCCGGGCT	ATACTTATAT	CAACCCTCTC	GACGGCACTT	ATCCGCCTGG	TACTGAGCAA	1980
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT	2040
2041	CAGAATAATA	GGTTCGGAAA	TAGGCAGGGG	GCATTAAC	TTTATACGGG	CACTGTTACT	2100
2101	CAAGGCACTG	ACCCCGTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG	2160
2161	TATGACGCTT	ACTGGAACGG	TAATTCAAGA	GA	CCATTCTGG	CTTTAATGAA	2220
2221	GATCCATTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCTGTCAAT	2280
2281	GCTGGCGGGCG	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGGTGGCTC	TGGTCCGGT	2400
2401	GATTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	2460
2461	AAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	2520
2521	GCTGCTATCG	ATGGGTTTAC	TGGTGACGT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT	2580
2581	GGTGATTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCAACCT	2640
2641	TTAATGAATA	ATTTCCGTCA	ATATTTCACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT	2700
2701	TTTGCTTTA	GCGCTGGTAA	ACCATATGAA	TTTTCTATTG	ATTGTGACAA	AATAAACTTA	2760
2761	TTCCGTGGTG	TCTTGTGCGT	TCTTTGCGT	GTTGCCACCT	TTATGTATGT	ATTTCTACG	2820
2821	TTTGCTAAC	TAATGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCGT	2880
2881	TATTATTGCG	TTTCTCTGGT	TTCTCTTCTGG	TAACCTTGTT	CGGCTATCTG	CTTACTTTTC	2940
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTCTT	GTTCTTGCT	CTTATTATTG	3000
3001	GGCTTAAC	AAATCTTG	GGTTATCTCT	CTGATATTAG	CGCTCAATT	CCCTCTGACT	3060
3061	TTGTTCAAGGG	TGTTCAAGTTA	ATTCCTCCGT	CTAATGCGCT	TCCCTGTTT	TATGTTATT	3120
3121	TCTCTGTA	GGCTGCTATT	TTCATTTTG	ACGTTAAACAA	AAAAATCGTT	TCTTATT	3180
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTTT	GTAACTGGCA	AATTAGGCTC	TGAAAGACG	3240
3241	CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT	3300
3301	CTTGATTTAA	GGCTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAC	GCCTCGCGTT	3360
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTGCTTG	CTATTGGGCG	CGGTAAATGAT	3420
3421	TCCTACGATG	AAAATAAAAA	CGGCTTGCTT	GTTCTCGATG	AGTGCAGGTAC	TTGGTTAAT	3480
3481	ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CGGATTATTG	ATTGGTTTCT	ACATGTCG	3540
3541	AAATTAGGAT	GGGATATTAT	TTTTCTTGTT	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG	3600
3601	CGTTCTGCAT	TAGCTGAACA	TGTTGTTTAT	TGTCGTGTC	TGGACAGAA	TACTTTACCT	3660
3661	TTTGTGGTGA	CTTATATTC	CTTATTACT	GGCTGAAAAA	TGCTCTG	TAAATTACAT	3720
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT	3780

3781	ACTGGTAAGA	ATTTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTTCTAG	TAATTATGAT	3840
3841	TCCGGTGT	ATTCTTATT	AACGCCATT	TTATCACACG	GTCGGTATT	CAAACCATT	3900
3901	AATTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATAATTG	AAAAGTTTC	ACGCCTTCT	3960
3961	TGTCTTGC	TTGGATTTC	ATCAGCATT	ACATATAGTT	ATATAACCCA	ACCTAAGCCG	4020
4021	GAGGTTAAAA	AGGTAGTCTC	TCAGACCTAT	GATTTTGATA	AATTCACTAT	TGACTCTTCT	4080
4081	CAGCGTCTT	ATCTAAGCTA	TCGCTATGTT	TTCAAGGATT	CTAAGGGAAA	ATTAATTAA	4140
4141	AGCGACGATT	TACAGAAGCA	AGGTTATTCA	CTCACATATA	TTGATTATG	TACTGTTTC	4200
4201	ATTAAGGAG	GTAATTCAA	TGAAATTGTT	AAATGTAATT	AATTTGTTT	TCTTGATGTT	4260
4261	TGTTTCATCA	TCTTCTTTG	CTCAGGTAAT	TGAAATGAAT	AATTGCGCTC	TGCGCGATT	4320
4321	TGTAACCTGG	TATTCAAAGC	AATCAGGCGA	ATCCGTTATT	GTTCCTCCG	ATGTAAGG	4380
4381	TACTGTTACT	GTATATTCA	CTGACGTTAA	ACCTGAAAAT	CTACGCAATT	TCTTATTTC	4440
4441	TGTTTACGT	GCTAATAATT	TTGATATGGT	TGGTTCAATT	CCTTCATAA	TTCAGAAGTA	4500
4501	TAATCCAAAC	AATCAGGATT	ATATTGATGA	ATTGCCATCA	TCTGATAATC	AGGAATATGA	4560
4561	TGATAATTCC	GCTCCTCTG	GTGGTTTCTT	TGTTCCGCAA	AATGATAATG	TTACTCAAAC	4620
4621	TTTAAATT	AATAACGTT	GGCAAGGAA	TTAATACGA	GTTGTCGAAT	TGTTGTAAA	4680
4681	GTCTAATACT	TCTAAATCCT	CAAATGTATT	ATCTATTGAC	GGCTCTAATC	TATTAGTTGT	4740
4741	TAGTGCACCT	AAAGATATT	TAGATAACCT	TCCTCAATT	CTTCTACTG	TTGATTTGCC	4800
4801	AACTGACCAG	ATATTGATTG	AGGGTTGAT	ATTTGAGGTT	CAGCAAGGTG	ATGCTTAA	4860
4861	TTTTCTATT	GCTGCTGGCT	CTCAGCGTGG	CACTGTTGCA	GGCGGTGTTA	ATACTGACCG	4920
4921	CCTCACCTCT	TTTTATCTT	CTGCTGGTGG	TTCGTTCGGT	ATTTTAATG	GCGATGTTT	4980
4981	AGGGCTATCA	TTTCAGGCT	TAAGACTAA	TAGCCATTCA	AAAATATTGT	CTGTGCCACG	5040
5041	TATTCTACG	GTGACTGGTGT	AAGAGGGTT	TATCTCTGTT	GGCCAGAATG	TCCCTTTTAT	5100
5101	TACTGGTCGT	GGTATTTC	AATCTGCCAA	TGTAATAAT	CCATTTCAGA	CGATTGAGCG	5160
5161	TCAAAATGTA	GGCAGCTTT	TGAGCGTTT	TCCTGTTGCA	ATGGCTGGCG	GTAATATTGT	5220
5221	TCTGGATATT	ACCAGCAAGG	CCGATAGTTT	GAGTTCTTCT	ACTCAGGCAA	GTGATGTTAT	5280
5281	TACTAATCAA	AGAAGTATTG	CTACAAACGGT	TAATTGCGT	GATGGACAGA	CTCTTTACT	5340
5341	CGGTGGCCTC	ACTGATTATA	AAAACACTTC	TCAAGATTCT	GGCGTACCGT	TCCTGCTAA	5400
5401	AATCCCTTA	ATCGGCCTCC	TGTTTAGCTC	CCGCTCTGAT	TCCAACGAGG	AAAGCACGTT	5460
5461	ATACGTGCTC	GTCAAAGCAA	CCATAGTAGC	CGCCCTGTAG	CGGCGCATT	AGCGCGGC	5520
5521	GTGTGGTGGT	TACGCGCAGC	GTGACCGCTA	CACTTGCAG	CGCCCTAGCG	CCCCTCCTT	5580
5581	TCGCTTCTT	CCCTTCTTT	CTGCCACGT	TCGCCGGCTT	TCCCCGTCAA	GCTCTAAATC	5640
5641	GGGGGCTCCC	TTTAGGGTTC	CGATTTAGTG	CTTACGGCA	CCTCGACCCC	AAAAAAACTG	5700
5701	ATTTGGGTGA	TGGTTCACGT	AGTGGGCCAT	CGCCCTGTAG	GACGGTTTT	CGCCCTTTGA	5760
5761	CGTTGGAGTC	CACGTTCTT	AAATAGTGGAC	TCTTGTCCA	AACTGGAACA	ACACTCAACC	5820
5821	CTATCTCGGG	CTATTCTTT	GATTATAAG	GGATTTCGAA	GATTTCGAA	CCACCATCAA	5880
5881	ACAGGATT	CGCCTGCTGG	GGCAACCCAG	CGTGGACCGC	TTGCTGCAAC	TCTCTCAGGG	5940
5941	CCAGGCCTG	AAGGGCAATC	AGCTGTTGCG	CGTCTCGCTG	GTGAAAGAA	AAACCACCT	6000
6001	GGCGCCCAAT	ACGAAACCCG	CCTCTCCCCG	CGCCTGGCC	GATTCAATTAA	TGAGCTGGC	6060
6061	ACGACAGGTT	TCCCAGCTGG	AAAGCGGGCA	GTGAGCGCAA	CGCAATTAA	GTGAGTTAGC	6120
6121	TCACTCATT	GGCACCCAG	GCTTACACT	TTATGTTCC	GGCTCGTATG	TTGTGTGAA	6180
6181	TTGTGAGCGG	ATAACAATT	CACACGCGTC	ACTTGGCACT	GGCCCTCGTT	TTACAACGTC	6240
6241	GTGACTGGG	AAACCCCTGGC	GTTACCCAAG	CTTGTACAT	GGAGAAAATA	AAGTGAACAA	6300
6301	AAGCACTATT	GCACCTGGC	TCTTACCGTT	ACCGTTACTG	TTTACCCCCG	TGACAAAAGC	6360
6361	CGCCCAAGTC	CAGCTGCTCG	AGTCAGGCCT	ATTGTGCCA	GGGGATTGTA	CTAGTGGATC	6420
6421	CTAGGCTGAA	GGCGATGACC	CTGCTAAGGC	TGCATTCAAT	AGTTTACAGG	CAAGTGCTAC	6480
6481	TGAGTACATT	GGCTACGCTT	GGGCTATGGT	AGTAGTTATA	GTTGGTCTA	CCATAGGGAT	6540
6541	TAAATTATT	AAAAAGTTA	CGAGCAAGGC	TTCTTAAAGCA	ATAGCGAAGA	GGCCCGCACC	6600
6601	GATGCCCTT	CCCAACAGTT	GCGCAGCTG	AATGGCGAAT	GGCGCTTGC	CTGGTTTCCG	6660
6661	GCACCAAGAAG	CGGTGCCGGA	AAGCTGGCTG	GAGTGCAGTC	TTCCCTGAGGC	CGATACGGTC	6720
6721	GTCGCTCCCT	CAAACCTGGCA	GATGCACGGT	TACGATGCGC	CCATCTACAC	CAACGTAAAC	6780
6781	TATCCCATTA	CGGTCAATCC	GCCGTTTGT	CCCACGGAGA	ATCCGACGGG	TTGTTACTCG	6840
6841	CTCACATT	ATGTTGATGA	AAGCTGGCTA	CAGGAAGGCC	AGACCGCGAAT	TATTTTGAT	6900
6901	GGCGTCCCTA	TTGGTTAAAA	AATGAGCTGA	TTAACAAAA	ATTTAACGCG	AATTAAACA	6960
6961	AAATATTAAAC	GTTTACAATT	AAATATTG	CTTATACAAT	CTTCTGTGTT	TTGGGGCTTT	7020
7021	TCTGATTATC	AACCAGGGTA	CATATGATTG	ACATGCTAGT	TTTACGATA	CCGTTCATCG	7080
7081	ATTCTTCTTGT	TTGCTCCAGA	CTCTCAGGCA	ATGACCTGAT	AGCCTTGT	GATCTCTAA	7140
7141	AAATAGCTAC	CCTCTCCGGC	ATTAATTAT	CAGCTAGAAC	GGTTGTT	CATATTGATG	7200
7201	GTGATTGAC	TGTCTCCGGC	CTTCTCACC	CTTTGAATC	TTACCTTCA	CATTACTCAG	7260
7261	GCATTGCTT	AAAAATATAT	GAGGGTTCTA	AAAATTTTA	TCCTTGGGTT	GAAATAAAGG	7320
7321	CTTCTCCCGC	AAAAGTATT	CAGGGTCATA	ATGTTTTGG	TACAACGGT	TTAGCTTAT	7380
7381	GCTCTGAGGC	TTTATTGCTT	AATTGGCTA	ATTCTTGCC	TTGCCGTT	GATTATTG	7440
7441	ACGTT						7445

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTCA	CTCGCGCCCC	AAATGAAAAT
61	ATAGCTAAC	AGGTTATTGA	CCATTGCGA	AATGTATCTA	ATGGTCAAC	AAATCTACT
121	CGTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAAGATT	AGCAATTAAAG	CTCTAAGCCA
241	TCCGAAAAAA	TGACCTCTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG
301	TTGGAGTTG	CTTCCGGTCT	GGTTCGCTT	GAAGCTCGAA	TTAAAAGCGG	ATATTGAAAG
361	TCTTCGGGC	TTCCCTCTAA	TCTTTTGAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGA	ACTTTAAAGCA
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT
541	AAACATTTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTG	CAAAGCCCTC	TCGCTATT
601	GGTTTTTATC	GTCGTCGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCTCGT
661	AATTCCCTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCTAA	ATCTCAACTG
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CGCTTAGTTC	GTTTTTATTAA	CGTAGATT
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTCTTA	AAATCGCATA	AGGTAATTCA
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGT
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC
1081	GTCTGCGCCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTG	CGGATTTCGA	CACAATT
1141	CAGGCGATGA	TACAAATCTC	CGTTGACTT	TGTTTCGCG	TTGGTATAAT	CGCTGGGGT
1201	CAAAGATGAG	TGTTTAGTG	TATTCTTCG	CCTCTTTCG	TTAGGTTGG	TGCCTTCGTA
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACTCTCTC	ATGAAAAAGT	CTTCTAGT
1321	CAAAGCCTCT	GTAGCCGGT	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA
1381	CGATCCCGCA	AAAGCGGCCT	TTAACCTCC	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA
1441	TGCGTGGGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GTTATCAAGC	TGTTTAAGAA
1501	ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTT	GGAGCCTTT
1561	TTTTGGAGA	TTTCAACGT	AAAAAAATTA	TTATTGCAA	TTCTTTAGT	TGTTCTTTC
1621	TATTCTCACT	CCGCTGAAC	TGTTGAAAGT	TGTTTAGCAA	AAACCCCATAC	AGAAAATTCA
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAAT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTG
1741	CTGTGGAATG	CTACAGGCGT	TGTTAGTTGT	ACTGGTGACG	AAACCTAGT	TTACGGTACA
1801	TGGGTTCTA	TTGGGCTG	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATAACACT
1921	ATTCCGGGCT	ATACTTAT	CAACCCTCTC	GACGGCACTT	ATCCGCTGG	TACTGAGCAA
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTT
2041	CAGAATAATA	GGTTCCGGAA	TAGGCAGGGG	GCATTAAC	TTTACGCGG	CACTGTTACT
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG
2161	TATGACGCTT	ACTGGAACCG	TAAATTCA	GA	ACTGCGCTT	TCCATTCTGG
2221	GATCCATTG	TTTGTGAATA	TCAAGGCCAA	TCGTCG	TGCGCTAAC	TCCGTCAAT
2281	GCTGGCGGCG	GCTCTGGTGG	TGGTCTG	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT
2341	GGCGGTTCTG	AGGGTGGCG	CTCTGAGGGG	GGCGGTTCCG	GTGGTGGCTC	TGGTCCGGT
2401	GATTTGATT	ATGAAAAGAT	GGCAACAGCT	AATAAGGGG	CTATGACCGA	AAATGCCGAT
2461	AAAAACGCGC	TACAGTCTG	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT
2521	GCTGCTATCG	ATGGTTTCA	TGGTACGTT	TCCGGCTTG	CTAATGGTAA	TGGTGCTACT
2581	GGTGATTTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACACT
2641	TTAATGAATA	ATTTCCGTC	ATATTACCT	TCCCTCC	AATCGGTTGA	ATGTCGCCCT
2701	TTTGTCTTTA	GCGCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTGACAA	AATAAACTTA
2761	TTCCGTGGTG	TCTTTCGCTT	TCTTTATAT	GTTGCCACCT	TTATGTTATG	ATTTCTACG
2821	TTTGCTAAC	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCGT
2881	TATTATTGCG	TTTCTCGGT	TTCCCTCTG	TAACTTGTT	CGGCTATCTG	CTTACTTTTC
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTCTT	GTTTCTGCT	CTTATTATTG
3001	GGCTTAAC	AATTCTTGTG	GGTTATCTC	CTGATATTAG	CGCTCAATT	CCCTCTGACT
3061	TTGTTCAAGGG	TGTTCAAGT	ATTCTCCC	CTAATGCGCT	TCCCTGTTT	TATGTTATT
3121	TCTCTGTAA	GGCTGCTATT	TTCA	ACGTTAAACA	AAAAATCGTT	TCTTATTG
3181	ATTGGGATAA	ATAATATG	TGTTTATTTT	TGAACTGGCA	AATTAGGCTC	TGAAAGACG
3241	CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGC	AGCAACTAAT
3301	TTTGATTAA	GGCTTCAAAA	CCTCCCGCAA	GTGCGGAGGT	TCGCTAAAC	GCCTCGCGT
3361	CTTAGAATAAC	CGGATAAGCC	TTCTATATCT	GATTTGCTG	CTATTGGGCG	CGTAATGAT
3421	TCCTACGATG	AAAAT	CGGCTGCTT	GTTC	CGT	3420
3481	ACCCGTTCTT	GGAAATGATAA	GGAAAGACAG	CGGATTATTG	AGTGC	3480
3541	AAATTAGGAT	GGGATATTAT	TTTTCTTGT	CAGGACTTAT	GGTGC	3540
3601	CGTTCTGCAT	TACGTGAAC	TGTTGTTTAT	CTATTGTTGA	AAACAGGCG	3600
3661	TTTGTGGT	CTTATATTC	TCTTATTACT	GGCTCGAAAA	TGCTCTG	TCTTACCT
3721	GTGGCGTT	TTAAATATG	CGATTCTCA	TAAAGCCCTA	CTGTC	3720
3781	CTCGGAA	TTTGTG	CGCTATG	CGTTC	GGCG	3780

3841	TCCGGTGT	TTT	ATTCTTATT	AACGCC	TTATCACACG	GT	CGGTAT	CAAACCA	3900
3901	AATTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATATTG	AAAAGTTTC	ACGCGT	TCT	3950	
3951	TGTCTTGC	TTGGATTG	ATCAGCATT	ACATATAG	ATATAACCC	ACCTAAG	CCG	4020	
4021	GAGGTTAAAA	AGGTAGTCTC	TCAGACCTAT	GATTTG	ATTCACTAT	TGACTCTT	TCT	4080	
4081	CAGCGTCT	ATCTAAGCTA	TCGCTATG	TC	CAAGGATT	CTAAGGGAAA	ATTAATTAA	4140	
4141	AGCGACGATT	TACAGAAGCA	AGGTTATTCA	CTCACATATA	TTGATTTATG	TACTGTT	TCC	4200	
4201	ATTAAGAAAG	GTAATTCAA	TGAAATTG	AAATGTAATT	AATTGTTG	TCTTGATG	TGTT	4260	
4261	TGTTCATCA	TCTTCTT	CTCAGGTAAT	TGAAATGAAT	ATTGCGCTC	TGCGCGAT	TT	4320	
4321	TGTAAC	TGG	TATTCAAAGC	AATCAGGCGA	ATCCGTTATT	TTTCTCCG	ATGTAAGG	4380	
4381	TACTGTTACT	GTATATTCA	CTGACGTTAA	ACCTGAAAT	CTACGCAATT	TCTTATTTC	4440		
4441	TGTTTACGT	GCTAATAATT	TTGATATGGT	TGGTCAATT	CCTTCATAA	TTCAGAAGTA	4500		
4501	TAATCCAAAC	AATCAGGATT	ATATTGATGA	ATTGCCATCA	TCTGATAATC	AGGAATATGA	4560		
4561	TGATAATTCC	GCTCCTTCTG	GTGGTTCTT	TGTTCCGCA	AATGATAATG	TTACTCAAC	4620		
4621	TTTAAATT	AATAACGTT	GGGCAAAGGA	TTAATACGA	GTTGTCGA	TGTTGTAA	4680		
4681	GTCTAACT	TCTAAATCCT	CAAATGTATT	ATCTATTGAC	GGCTCTAA	TATTAGTTG	4740		
4741	TAGTGCACCT	AAAGATATT	TAGATAACCT	TCCTCAATT	CTTCTACTG	TTGATTG	GCC	4800	
4801	AACTGACCA	ATATTGATTG	AGGGTTGAT	ATTGAGGTT	CAGCAAGGTG	ATGCTT	AGA	4860	
4861	TTTTCA	GCTGCTG	CTCAGCGT	CACTG	GGCGGTG	ATACTGACCG	4920		
4921	CCTCACCTCT	GTTTATCTT	CTGCTGGT	TTGTTCGG	ATTTTAATG	GCGATG	TTT	4980	
4981	AGGGCTATCA	GTCGCGCAT	TAAAGACTAA	TAGCCATTCA	AAAATATTG	CTG	TGCCACG	5040	
5041	TATTCTTACG	CTTCAAGGTC	AGAAGGGTTC	TATCTCTG	GGCCAGAATG	TCCCTT	TAT	5100	
5101	TACTGGTC	GTGACTG	AATCTGCCA	TGTAATAAT	CCATTTCAGA	CGATTGAGC	G	5160	
5161	TCAAAATGTA	GGTATT	CCCA	TGAGCG	ATGGCTGG	GTAATT	TGT	5220	
5221	TCTGGTAT	ACCAGCAAGG	CCGATAGTT	GAGTTCTT	ACTCAGGCAA	GTGATGTT	TAT	5280	
5281	TACTAATCAA	AGAAGTATTG	CTACAA	CGGT	TAATTGCG	GATGGACAGA	CTCTTT	5340	
5341	CGGTGGCCTC	ACTGATTATA	AAAACACTTC	TCAAGATT	GGCGTACCG	TCC	TGTCTAA	5400	
5401	AATCCCTTA	ATCGGCC	TGTTAGCTC	CCGCTCTG	TCCAACGAGG	AAAGCACG	TT	5460	
5461	ATACGTGCTC	GTC	AAAGC	CCATAGTACG	CGCCCTG	GGCGCATT	AGCGCGG	5520	
5521	GTGTGGTGGT	TACGCGC	GTGACCGC	CACTTGC	CGCCCTAGCG	CCC	GTCCTT	5580	
5581	TCGCTTCTT	CCCTCCTT	CTCGCCACG	TCGCGG	TCCCCGT	GCTCTAA	ATC	5640	
5641	GGGGGCTCCC	TTT	AGGGT	CGATTAGT	CTTACGG	CCTCGAC	AAAAAA	5700	
5701	ATTTGGGTGA	TGGTTACG	AGTGGGCC	CGCCCTG	GACGGTTT	CGCC	TTGA	5760	
5761	CGTTGGAGTC	CACGTTCTT	AATAGTGGAC	TCTTGT	AACTGGAA	ACACTCAAC	5820		
5821	CTATCTCGGG	CTATTCTT	GATTATAAG	GGATT	GGATTTCGG	CCACCATCA	5880		
5881	ACAGGATT	CGCCTG	GGCAAACCA	CGTGGACCG	TGCTGCA	TCTCTCAGGG	5940		
5941	CCAGGGCGGT	AAAGGCA	AGCTGTTG	CGTCTCG	GTGAAAAG	AAACCAACCT	6000		
6001	GGCGCCCA	ACGCAAACCG	CCTCTCCC	CGCGTTGG	GATTCTTAA	TGCAGCTGG	6060		
6061	ACGACAGGTT	TC	AAAGCGGGCA	GTGAGCG	CGCAATTAA	GTGAGTTAGC	6120		
6121	TCACTCATTA	GGCACCC	GCTTAC	TTATGCTT	GGCTCGT	TTGTGTG	6180		
6181	TTGTGAGCGG	ATAACA	TTACACG	GGAGACAG	ATAATGAA	ACCTATTG	6240		
6241	TACGGCAGCC	GCTGGATTG	TATTACT	TGCCC	GCATGGCC	AGCTCGT	GAT	6300	
6301	GACCCAGACT	CCAGATA	AACAGGA	AGTGT	CTAGAACG	TCACTGG	6360		
6361	CTGGCCGTC	TTT	ACG	GATG	GCGTACCC	AGCTTAATG	6420		
6421	CCTTGCAGAA	CC	CCAGCTGG	TAATAG	GAGGCCG	CCGATCG	6480		
6481	TTCCCAACAG	TTGCGCAG	TGATTGG	ATGGC	GCCTGG	GGCACCA	6540		
6541	AGCGGTGCCG	GAAAGCTG	TGGAGT	TCTTC	GCGATACG	TCGTGT	6600		
6601	CTAAACTGG	CAGATG	GTTACG	GCCC	ACCAACG	CCTATCCC	6660		
6651	TACGGTCA	CCGCG	TTG	GAATCCG	GGTTGTT	CGCTCAC	6720		
6721	TAATGTTGAT	GAAAGCTG	TACAGGA	GAAGC	ATTTTTT	ATGGCGT	6780		
6781	TATTGGTTAA	AAAATGAG	GATTAA	AAATT	CGAATT	CAAAATATT	6840		
5841	ACGTTTACAA	TTTAAAT	TGCTTATA	ATCTT	TTTGGG	TTTCTGAT	6900		
6901	TCAACCGGGG	TACATATG	TGACATG	GTTT	TACG	CGATTCT	6960		
6961	GTTTGC	GA	CAATGAC	GAATG	TACG	TGATG	7020		
7021	ACCCCTCTCCG	GC	ATCAG	ACGG	TCTCT	AAAAAATAGC	7080		
7081	ACTGTCTCCG	GC	CTAGA	TTG	CTAC	TGGTGATT	7140		
7141	TTAAATAT	GTGAGG	CC	TCTT	ACAT	AGGCATTG	7200		
7201	GC	TAATG	GGTAC	GGTAC	TTG	GGCTTCTC	7260		
7261	GCTTATTG	TA	TTT	TTG	GGAT	ATGCTCTG	7320		
	10	20	30	40	50	60	70		

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTCAAG	CTCGCGCCCC	AAATGAAAAT
61	ATAGCTAAC	AGGTTATTGA	CCATTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT
121	CGTTCGAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAAGATT	AGCAATTAAG	CTCTAAGCCA
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG
301	TTGGAGTTG	CTTCCGGTCT	GGTTGCTTT	GAAGGTCGAA	TTAAAACGCG	ATATTGAAAG
361	TCTTCGGGC	TTCCCTTTAA	TCTTTTGT	GCAATCCGCT	TTGCTCTGA	CTATAATAGT
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGA	GTTAAAGCA
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT
541	AAACATTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTG	CAAAGCCTC	TCGCTATTTT
601	GGTTTTATC	GTCGTCGTT	AAACGAGGGT	TATGATAGT	TTACGTTGAT	TTGGGTAATG
661	AATTCTTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCTAA	ATCTCAACTG
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATT
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGT
901	CTCGTCAGGG	CAAGCCCTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC
1081	GTCTGCGCCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTGCG	CGGAGTTCGA	CACAATTAT
1141	CAGGCATGA	TACAAATCTC	CGTTGACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT
1201	CAAAGATGAG	TGTTTTAGTG	TATTCTTCG	CCTCTTCG	TTTAGGTTGG	TGCCTTCGTA
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACTCCCTC	ATGAAAAAGT	CTTCTAGTCCT
1321	CAAAGCCTCT	GTAGCCGTTG	CTAACCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA
1381	CGATCCCGCA	AAAGCGCCT	TAAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA
1441	TGCGTGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA
1501	ATTACACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTTT
1561	TTTTGGAGA	TTTCAACG	GAAAAAATTA	TTATTGCGAA	TTCCTTCTAGT	TGTTCCCTTC
1621	TATTCTCACT	CCGCTGAAAC	TGTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA
1681	TTTACTAACG	TCTGAAAGA	CGACAAACACT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT
1741	CTGTGGAATG	CTACAGGC	TGTAGTTGTT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA
1801	TGGGTTCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT
1921	ATTCCGGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACTT	ATCCGCCCTG	TACTGAGCAA
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT
2041	CAGAATAATA	GGTTCCGAAA	TAGGCAGGGG	GCATTAAC	TTTATACGGG	CACTGTTACT
2101	CAAGGCAC	ACCCCGTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG
2161	TATGACGCTT	ACTGGAACGG	AAATTCA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA
2221	GATCATTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT
2281	GCTGGCGCG	GCTCTGGT	TGGTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGGTGGCTC	TGGTCCGGT
2401	GATTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT
2461	AAAACCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGT	CTGTCTGCTAC	TGATTACGGT
2521	GCTGCTATCG	ATGGTTTCA	TGGTACGTT	TCCGGCTTG	CTAATGGTAA	TGGTGTACT
2581	GGTGA	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACCT
2641	TTAATGAATA	ATTTCCGTC	ATATTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTGCCCC
2701	TTTGTCTTTA	GCGCTGGTAA	ACCATATGAA	TTTTCTATTG	ATTGTGACAA	AATAAACTTA
2761	TTCCGTGGT	TCTTTCGTT	TCTTTTATAT	GTTGCCACCT	TTATGTATG	ATTTCTACG
2821	TTTGCTAAC	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCGT
2881	TATTATTGCG	TTTCCTCGGT	TTCCCTCTGG	TAACCTTGT	GGCGTATCTG	CTTACTTTTC
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTCTT	GTTCCTTGCT	CTTATTATTG
3001	GGCTTAAC	AATTCTGTG	GGTTATCTCT	CTGATATTAG	CGCTCAATT	CCCTCTGACT
3061	TTGTCAGGG	TGTTCAAGTTA	ATTCTCCGT	CTAATGCGCT	TCCCTGTTT	TATGTTATT
3121	TCTCTGTA	GGCTGCTATT	TTCACTTTTG	ACGTTAAACA	AAAAATCGTT	TCTTATTG
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTTT	GTAACTGGCA	AATTAGGCTC	TGAAAGACG
3241	CTCGTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTCAAAAT	AGCAACTAAT
3301	CTTGATTAA	GGCTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAAC	GCCTCGCGTT
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTGCTT	CTATTGGCG	CGGTAATGAT
3421	TCCTACGATG	AAAATAAAAA	CGGCTTGCTT	GTTCTCGATG	AGTGCAGGTC	TTGGTTAAT
3481	ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CCGATTATTG	ATTGGTTCT	ACATGCTCGT
3541	AAATTAGGAT	GGGATATTAT	TTTTCTGTT	CAGGACTTAT	CTATTGTTGA	TAACACGGCG
3601	CGTTCTGCA	TAGCTGAACA	TGTTGTTTAT	TGTCGTCGTC	TGGACAGAAT	TACTTTACCT
3661	TTTGTGCGTA	CTTATATTC	TCTTATTACT	GGCTCGAAAA	TGCTCTGCG	TAATTACAT
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTAA	TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT
3781	ACTGGTAAGA	ATTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT

7 / 11

3841	TCCGGTGT	TTT	ATTCTTATT	AACGCCTT	AT	TTATCACACG	GTCGGTATT	CAAACCATTA	3900
3901	AATTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATATTG	AAAAGTTTC	ACGGGTTCT	3960		
3961	TGTCTTGC	GA	TTGGATTTC	ATCAGCATT	ACATATAGTT	ATATAACCCA	ACCTAAGCCG	4020	
4021	GAGGTAAAAA	AGGTAGTCTC	TCAGACCTAT	GATTTGATA	AATTCACTAT	TGACTCTTCT	4080		
4081	CAGCGTCTT	ATCTAAGCT	TCGCTATGTT	TTCAAGGATT	CTAAGGGAAA	ATTAATTAAT	4140		
4141	AGCGACGATT	TACAGAAC	AGGTTATTCA	CTCACATATA	TTGATTATG	TACTGTTCC	4200		
4201	ATTAAGGAAAG	GTAATTCAA	TGAAATTGTT	AAATGTAATT	AATTTGTTT	TCTTGATGTT	4260		
4261	TGTTTCA	TCTTCTT	TCAGGTAAT	TGAAATGAAT	AATTGCCTC	TGCGCGATT	4320		
4321	TGTAAC	TGG	TATTCAAAGC	AATCAGGCGA	ATCCGTTATT	GTTCCTCCG	ATGTAAGG	4380	
4381	TACTGTTACT	GTATATT	CTGACGTTAA	ACCTGAAAT	CTACGCAATT	TCTTATTTC	4440		
4441	TGTTTACGT	GCTAATAATT	TTGATATGGT	TGGTTCAATT	CCTTCCATAA	TTCAGAAGTA	4500		
4501	TAATCCAAAC	AATCAGGATT	ATATGATGA	ATTGCCATCA	TCTGATAATC	AGGAATATGA	4560		
4561	TGATAATTCC	GCTCCTT	GTGGTTCTT	TGTTCCGAA	AATGATAATG	TTACTCAAAC	4620		
4621	TTTTAAATT	AATAACGTT	GGGCAAAGGA	TTAATACGA	TTGTCGAAT	TGTTGTAAA	4680		
4681	GTCTAAT	TCTAAATCCT	CAAATGTT	ATCTATTGAC	GGCTCTAATC	TATTAGTTGT	4740		
4741	TAGTGAC	AAAGATATT	TAGATAACCT	TCCTCAATT	CTTCTACTG	TTGATTGCC	4800		
4801	AACTGACCAG	ATATTGATT	AGGGTTGAT	ATTTGAGGTT	CAGCAAGGTG	ATGCTT	4860		
4861	TTTTCA	TCTGCTGG	CTCAGCGTGG	CACTGTTGCA	GGCGGTGTTA	ATACTGACCG	4920		
4921	CCTCAC	CT	GTTCCTGTT	TTCGTTGGT	TTCGTTGGT	ATTTTAATG	GCAGATGTTT	4980	
4981	AGGGCTATCA	GTTCGCGCAT	TAAAGACTAA	TAGCCATTCA	AAAATATTGT	CTGTGCCACG	5040		
5041	TATTCTTACG	CTTTCAGGTC	AGAAGGGTTC	TATCTCTGTT	GGCCAGAATG	TCCCTTTTAT	5100		
5101	TACTGGTCGT	GTGACTGGT	AATCTGCCAA	TGTAATAAT	CCATTTCAGA	CGATTGAGCG	5160		
5161	TCAAAATGTA	GGTATTTC	TGAGCGTTT	TCCTGTTGCA	ATGGCTGGCG	GTAATATTGT	5220		
5221	TCTGGATATT	ACCAGCAAGG	CCGATAGTT	GAGTTCTT	ACTCAGGCAA	GTGATGTT	5280		
5281	TACTAATCAA	AGAAGTATTG	CTACAAACGGT	TAATTGCGT	GATGGACAGA	CTCTTTACT	5340		
5341	CGGTGGCCTC	ACTGATTATA	AAAACACTTC	TCAAGATTCT	GGCGTACCGT	TCCTGTC	5400		
5401	AATCCCTTA	ATCGGCCTC	TGTTAGCTC	CCGCTCTGAT	TCCAACGAGG	AAAGCACGTT	5460		
5461	ATACGTGCTC	GTCAAAGCAA	CCATAGTACG	CGCCCTGTAG	CGGCCGCA	AGCGCGGCGG	5520		
5521	GTGTGGTGGT	TACGCGCAGC	GTGACCGCTA	CACTTGCAG	CGCCCTAGCG	CCCGCTCCTT	5580		
5581	TCGCTT	CCCTT	CTCGCCACGT	TCGCGGGCT	TCCCCGTCAA	GCTCTAAATC	5640		
5641	GGGGGCTCCC	TTTAGGGT	CGATTAGTG	CTTACGGCA	CCTCGACCCC	AAAAAACTTG	5700		
5701	ATTGGGTGA	TGGTTACGT	AGTGGGCCAT	CGCCCTGATA	GACGGTTTT	CGCCCTTGA	5760		
5761	CGTTGGAGTC	CACGTTCTT	AATAGTGGAC	TCTTGT	AACTGGAACA	ACACTCAACC	5820		
5821	CTATCTCGGG	CTATTCTT	GATTATAAG	GGATTG	GATTTGCGGA	CCACCATCAA	5880		
5881	ACAGGATT	CGCCTGCTG	GGCAAACCAG	CGTGGACCG	TTGCTGCAAC	TCTCTAGGG	5940		
5941	CCAGGGGGT	AAGGGCAATC	AGCTGTTGCC	CGTCTCGCTG	GTGAAAAGAA	AAACCAACCT	6000		
6001	GGGCCCAAT	ACGAAACCG	CCTCTCCCCG	CGCGTTGGCC	GATTCA	TGCACTG	6060		
6061	ACGACAGGTT	TCCC	AAAGCAGGCA	GTGAGCGCAA	CGCAATTAA	6120			
6121	TCACTCATTA	GGCACCCCAG	GCTTACACT	TTATGCT	GGCTCGTATG	TTGTGTGAA	6180		
6181	TTGTGAGCGG	ATAACAATT	CACACGCGTC	ACTTGGCACT	GGCCGTCGTT	TTACAACGTC	6240		
6241	GTGACTGGG	AAACCTGGC	GTTACCAAG	CTTGTACAT	GGAGAAAATA	AAGTGAACAA	6300		
6301	AAGCACTATT	GCAC	TCTTACCGTT	ACTGTTTAC	CCTGTTGCAA	AAGCCCAGGT	6360		
6361	CCAGCTGCTC	GAGTCGGT	TCCCCCTGGC	ACCCCTCTC	AAGAGCACCT	CTGGGGCAC	6420		
6421	AGCGGCCCTG	GGCTGCCTG	TCAAGACTAA	TTCCCCGAAAC	CGGTGACGGT	GTGTTGGAAC	6480		
6481	TCAGGGGCC	TGAC	CGTGCACACC	TTCCCGGCTG	TCCTACAGTC	CTCAGGACTC	6540		
6541	TACTCCCTCA	GCAGCGTGGT	GACCGTGCCC	TCCAGCAGCT	TGGGCACCC	GACCTACATC	6600		
6601	TGCAACGTGA	ATCACAAGCC	CAGCAACACC	AAGGTGGACA	AGAAAGCAGA	GCCCCAAATCT	6660		
6661	TGTACTAGT	GATCCTACCC	GTACGACGTT	CCGGACTACG	CTTCTTAGGC	TGAAGGCGAT	6720		
6721	GACCTGCTA	AGGCTG	CAATAGTTA	CAGGCAAGTG	CTACTGAGTA	CATTGGCTAC	6780		
6781	GCTTGGGCTA	TGGTAGTAGT	TATAGTTG	GCTACCATAG	GGATTAAATT	ATTCAAAAG	6840		
6841	TTTACGAGCA	AGGCTTCTT	AGCAATAGCG	AAGAGGCCG	CACCGATCG	CCTTCCAAAC	6900		
6901	AGTTGCGCAG	CCTGAATGGC	GAATGGCGCT	TTGCCTGGTT	TCCGGCACCA	GAAGCGGTGC	6960		
6961	CGGAAAGCTG	GCTGGAGTGC	GATCTTCTG	AGGCCGATAC	GGTCGTCGTC	CCCTCAAAC	7020		
7021	GGCAGATGCA	CGGTTACGAT	GCGCCCATCT	ACACCAACGT	AACCTATCCC	ATTACGGTCA	7080		
7081	ATCCGCC	GT	TGTTCCCACG	GAGAATCCGA	CGGGTTGTTA	CTCGCTCAC	TTAATGTT	7140	
7141	ATGAAAGCTG	GCTACAGGAA	GGCCAGACGC	GAATTATT	TGATGGCGTT	CCTATTGTT	7200		
7201	AAAAAAATGAG	CTGATT	AA	CGCGAATT	AACAAATAT	TAACGTTAC	7260		
7261	AATTAAATA	TTTGCT	TTA	AAAAATTAA	CGCGAATT	TAACGTTAC	7320		
7321	GGTACATATG	ATTGACATG	TACTTT	GGTTTTGGGG	CTTTTCTGAT	TATCAACC	7380		
7381	CAGACTCTCA	GGCAATGACC	TGATAGCCT	ATTACCGTTC	ATCGATTCTC	TTGTTGCTC	7440		
7441	CGGCATTAAAT	TTACAGCTA	GAACGGTTGA	ATATCATATT	GATGGTGATT	TGACTGTC	7500		
7501	CGGCCTTCT	CACCC	TTTG	AATCTTAC	TACACATTAC	TCAGGCATTG	7560		
7561	ATATGAGG	TCTAAAATT	TTTATCCTG	CGTTGAAATA	AAGGCTTCTC	CCGCAAAAGT	7620		
7621	ATTACAGGGT	CATAATGTT	TTGGTACAAC	CGATTAGCT	TTATGCTG	AGGCTTATT	7680		
7681	GCTTAATT	TT	TGCGCTTG	GTATGATT	TTGGACGTT	7720			

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTCA	CTCGCGCCCC	AAATGAAAAT
61	ATAGCTAAAC	AGGTTATTGA	CCATTTGCCA	AATGATCTA	ATGGTCAAAC	TAATCTACT
121	CGTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA
181	GTTCATATT	TAAAACATGT	TGAGCTACAG	CACCAAGATT	AGCAATTAAAG	CTCTAAGCCA
241	TCCGCAAAAA	TGACCTCTTA	TCAAAAGGGAG	CAATTAAAGG	TACTCTCTAA	TCTTGACCTG
301	TTGGAGTTG	CTTCCGGTCT	GGTTCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTGAAAG
361	TCTTCGGGC	TTCTCTTTAA	TCTTTTTGAT	GCAATCCGCT	TTGCTCTGA	CTATAATAGT
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGAAC	GTAAAGCA
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT
541	AAACATTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTG	CAAAAGCCCTC	TCGCTATTTT
601	GGTTTTATC	GTCGCTGTT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCCTCGT
661	AATTCCCTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCCTAA	ATCTCAACTG
721	ATGAATCTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTT	GTTTTATTAA	CGTAGATTTT
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAAATTCA
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGT
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCAT	GCGCCCTGGTC
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC
1081	GTCGCGCCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTG	CGGATTTCGA	CACAATTAT
1141	CAGGCATGTA	TACAAATCTC	CGTTGACTT	TGTTTGC	TTGGTATAAT	CGCTGGGGGT
1201	CAAAGATGAG	TGTTTGTG	TATTCTTCG	CCTCTTCG	TTAGGTTGG	TGCTTCGTA
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACTCCCTC	ATGAAAAAAGT	CTTTAGTCCT
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCCTCG	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA
1381	CGATCCCAGA	AAAGCGGCCT	TTAACCTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA
1441	TGCGTGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTAAGAA
1501	ATTACACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCCTTT
1561	TTTTGGAGA	TTTCAACGT	AAAAAAATTA	TTATTGCAA	TTCTTTAGT	TGTTCTTTTC
1621	TATTCTACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAAC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT
1741	CTGTGGAATG	CTACAGGC	TGTTAGTTGT	ACTGGTGACG	AAACTCAGT	TTACGGTACA
1801	TGGGTTCTA	TTGGGCTTC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGGCCTGGT
1861	TCTGAGGGTG	GCGGTTCTG	TGTTCTGAA	GTAGCTCAGC	CTCTTAATAC	TTTCATGTTT
1921	ATTCCGGGCT	ATACTTATAT	CAACCCCTTC	GACGGCACTT	CTGAGTACGG	TGATACACCT
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	ATCCGCCCTG	TACTGAGCAA
2041	CAGAATAATA	GGTCCGAAA	TAGGCAGGGG	GCATTAAC	CTCTTAATAC	TTTCATGTTT
2101	CAAGGCACTG	ACCCCCTAA	AACCTATTAC	CACTACACTC	TTTATACGGG	CACTGTTACT
2161	TATGACGCTT	ACTGGAACGG	TAAATTCTGAG	GA	CTGTATCATC	AAAAGCCATG
2221	GATCCATTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TCCATTCTGG	CTTTAATGAA
2281	GCTGGCGGCG	GCTCTGGT	TGGTCTGGT	TG	TGCTCAACC	TCCTGTCAT
2341	GGCGGTTCTG	AGGGTGGCG	CTCTGAGGGG	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT
2401	GATTTTGATT	ATGAAAAGAT	GGCAAAACGCT	GGCGGTTCCG	GTGGTGGCTC	TGGTCCGGT
2461	GAAAACGCGC	TACAGTCGA	CGCTAAAGGC	AATAAGGGGG	CTATGACCGA	AAATGCCGAT
2521	GCTGCTATCG	ATGGTTTCA	TGGTGA	AAACTTGATT	CTGTCGCTAC	TGATTACGGT
2581	GGTGA	CTGGCTCTAA	TTCCCAAATG	TCCGGCTTG	CTAATGGTAA	TGGTCTACT
2641	TTAATGAATA	ATTTCCGTC	ATATTACCT	GCTCAAGTCG	GTGACGGTGA	TAATTACACCT
2701	TTTGCTTTA	GCGCTGGTAA	ACCATATGAA	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT
2761	TTCCGTGGT	TCTTTCGTT	TCTTTATAT	TTTCTATTG	ATTGTGACAA	AATAAACCTA
2821	TTTGCTAAC	TACTGCGTAA	TAAGGAGTCT	GTTGCCAC	TTATGTATGT	ATTTCTACG
2881	TATTATTGCG	TTTCTCTGGT	TTCTTCTGG	TAATCATGCC	AGTTCTTTTG	GGTATTCCGT
2941	TTAAAAAGGG	CTTCGGTAA	ATAGCTATTG	TAACCTTGT	CGCTATCTG	CTTACTTTTC
3001	CTCAATTCTT	GTGGGTTATC	TCTCTGATAT	CCTGTTCTT	GCTCTTATT	TTGGGCTTAA
3061	GGGTGTTCA	TTAATTCTCC	CGCTAATGC	TAGCGCTCAA	TTACCCCTG	ACTTTGTTCA
3121	AAAGGCTGCT	ATTTTCATTT	TTGACGTTAA	GCTTCCCTGT	TTTTATGTTA	TTCTCTCTGT
3181	TAAATAATAT	GGCTGTTTAT	TTGTAAC	ACAAAAAAATC	GTTCTTATT	TGGATTGGGA
3241	GCGTGGTAA	GATTCAAGGAT	AAAATTGTA	GCAAAATTAGG	CTCTGGAAAG	ACGCTCGTTA
3301	TAAGGCTTC	AAACCTCCCG	CAAGTCGGGA	CTGGGTGCAA	AATAGCAACT	AATCTTGATT
3361	TACCGGATAA	GCCTTCTATA	TCTGATTTC	GGTTCGCTAA	AACGCCCTCGC	GTTCTTAGAA
3421	ATGAAAATAA	AAACGGCTTG	CTTGTTC	TTGCTATTGG	GCGCGGTTAAT	GATTCTACG
3481	CTTGGAAATG	TAAGGAAAGA	CAGCCGATT	ATGAGTCGG	TACCTGGTT	AATAACCGTT
3541	GATGGGATAT	TATTTTCTT	GTTCAAGGACT	TTGATTGGTT	TCTACATGCT	CGTAAATTAG
3601	CATTAGCTGA	ACATGTTGTT	TATTGTC	TATCTATTGT	TGATAAACAG	GCGCGTCTG
3661	GTACTTTATA	TTCTCTTATT	ACTGGCTCGA	GTCTGGACAG	AATTACTTTA	CCTTTGTG
3721	TTGTTAAATA	TGGCGATTCT	CAATTAAGCC	AAATGCCCT	GCCTAAATTAA	CATGGTGGCG
3781	AGAATTGTA	TAACCGCATT	GATACTAAAC	CTACTGTTG	GGGTTGGCTT	TATACGGTAA

3841	TTTATTCTTA	TTTAACGCCT	TATTTATCAC	ACGGTCGGTA	TTTCAAACCA	TTAAATTTAG	3900
3901	GTCAGAAGAT	GAAGCTACT	AAAATATATT	TGAAAAGTT	TTCACCGCGTT	CTTTGTCTTG	3960
3961	CGATTGGATT	TGCATCAGCA	TTTACATATA	GTTATATAAC	CCAACCTAAG	CCGGAGGTTA	4020
4021	AAAAGGTAGT	CTCTCAGACC	TATGATTTG	ATAAATTCAC	TATTGACTCT	TCTCAGCGTC	4080
4081	TTAATCTAAG	CTATCGCTAT	GTTCAGG	ATTCTAAGGG	AAAATTAATT	AATAGCGACG	4140
4141	ATTACAGAA	GCAAGGTTAT	TCACTCACAT	ATATTGATTT	ATGTACTGTT	TCCATTAAAA	4200
4201	AAGGTAATT	AAATGAAATT	GTAAATGTA	ATTAATTTG	TTTCTTGAT	GTTGTTTCA	4260
4261	TCATCTTCTT	TTGCTCAGGT	AATTGAAATG	AATAATTCGC	CTCTGCGCGA	TTTGTAACT	4320
4321	TGGTATTCAA	AGCAATCAGG	CGAATCCGTT	ATTGTTTCTC	CCGATGTAAG	AGGTACTGTT	4380
4381	ACTGTATATT	CATCTGACGT	TAACACCTGAA	AATCTACGCA	ATTTCTTAT	TTCTGTTTA	4440
4441	CGTCTAATA	ATTTTGATAT	GGTTGGTTCA	ATTCCTTCCA	TAATTCAAGAA	GTATAATCCZ	4500
4501	AACAATCAGG	ATTATAATTGA	TGAATTGCCA	TCATCTGATA	ATCAGGAATA	TGATGATAAT	4560
4561	TCCGCTCCTT	CTGGTGGTTT	CTTGTGTTCCG	CAAAATGATA	ATGTTACTCA	AACCTTTAAA	4620
4621	ATTAATAACG	TCGGGGCAAA	GGATTTAATA	CGAGTTGTCG	AATTGTTGT	AAAGTCTAAT	4680
4681	ACTTCTAAAT	CCTCAAATGT	ATTATCTATT	GACGGCTCTA	ATCTATTAGT	TGTTAGTGCA	4740
4741	CCTAAAGATA	TTTAGATAA	CCCTCCTCA	TTCCCTTCTA	CTGTTGATT	GCCAACTGAC	4800
4801	CAGATATTGA	TTGAGGGTTT	GATATTGAG	GTTCAGCAAG	GTGATGCTT	AGATTTTCA	4860
4861	TTTGCTGCTG	GCTCTCAGCG	TGGCACTGTT	GCAGGCGGTG	TTAAATACTGA	CCGCGCTCACC	4920
4921	TCTGTTTTAT	CTTCTGCTGG	TGTTCGTTC	GGTATTTTTA	ATGGCGATGT	TTTAGGGCTA	4980
4981	TCAGTCGCG	CATTAAGAGAC	TAATAGCCAT	TCAAAAATAT	TGTCTGTGCC	ACGTATTCTT	5040
5041	ACGCTTCAG	GTCAGAAGGG	TTCTATCTCT	GTGAGGCCAGA	ATGTCCCTT	TATTACTGGT	5100
5101	CGTGTGACTG	GTGAATCTGC	CAATGTAAT	AATCCATTTC	AGACGATTGA	GCGTCAAAAT	5160
5161	GTAGGTATTT	CCATGAGCGT	TTTCCCTGTT	GCAATGGCTG	GCGGTAAAT	TGTTCTGGAT	5220
5221	ATTACCAGCA	AGGCCGATAG	TTTGAGTTCT	TCTACTCAGG	CAAGTGATGT	TATTACTAAT	5280
5281	CAAAGAAGTA	TTGCTACAAAC	GGTTAATTG	CGTGATGGAC	AGACTCTTT	ACTCGGTGGC	5340
5341	CTCACTGATT	ATAAAAAACAC	TTCTCAAGAT	TCTGGCGTAC	CGTTCTGTC	AAAAATCCCT	5400
5401	TTAATCGGCC	TCCTGTTAG	CTCCCGCTCT	GATTCAACG	AGGAAAGCAC	GTTATACGTG	5460
5461	CTCGTCAAAG	CAACCATAGT	ACGCGCCCTG	TAGCGGCGCA	TTAAGCGCGG	CGGGTGTGGT	5520
5521	GGTACGCGC	AGCGTACCG	CTACACTTGC	CAGCGCCCTA	GCGCCCGCTC	CTTCGCTTT	5580
5581	TTTCGCCTGC	TGGGGCAAC	CAGCGTGGAC	CGCTTGTGC	AACTCTCTA	GGGCCAGGCG	5940
5941	GTGAAGGGCA	ATCAGCTGTT	GCCCCTCTCG	CTGGTAAAAA	GAAAAAACAC	CCTGGCGCCC	6000
6001	AATACGCAAA	CCGCCTCTCC	CCGCGCCTTG	GGCGATTCA	TAATGCAGCT	GGCACGACAG	6060
6061	GTTCCTCGAC	TGGAAAGCGG	GCAGTGAGCG	CAACGCAATT	AATGTGAGTT	AGCTCACTCA	6120
6121	TTAGGCACCC	CAGGCTTAC	ACTTTATGCT	TCCGGCTCGT	ATGTTGTGT	GAATTGTGAG	6180
6181	CGGATAACAA	TTTCACACGC	CAAGGAGACA	GTCATAATGA	AATACCTATT	GCCTACGGCA	6240
6241	GCCGCTGGAT	TGTTATTACT	CGCTGCCAA	CCAGCCATGG	CCCGACATCT	CCCGCCATCT	6300
6301	GATGAGCAGT	TGAAATCTGG	AACTGCCCTC	GTTGTGTGCC	TGCTGAATTA	CTTCTATCCC	6360
6361	AGAGAGGCCA	AAGTACAGTG	GAAGGGTGGAT	AAACGCCCTC	AAATGGGTAA	CTCCCAAGGAG	6420
6421	AGTGTACAG	AGCAGGACAG	CAAGGACAGC	ACCTACAGCC	TCAGCAGCAC	CCTGACGCTG	6480
6481	AGCAAAGCAG	ACTACGAGAA	ACACAAAGTC	TACGCTGCC	AAGTCACCCA	TCAGGGCTG	6540
6541	AGCTCGCCCG	TCACAAAGAG	CTTCAACAGG	GGAGAGTGT	CTAGAACGCG	TCACTTGGCA	6600
6601	CTGGCCGTG	TTTACAACG	TCGTGACTGG	GAAAACCTG	GCGTTACCCA	AGCTTAATCG	6660
6661	CCTTCAGAA	TTCCCTTTG	CCAGCTGGCG	TAATAGCGAA	GAGGCCCGCA	CCGATCGCCC	6720
6721	TTCCCACAG	TTGCGCAGCC	TGAATGGCGA	ATGGCGCTT	GCCTGGTT	CGGACCCAGA	6780
6781	AGCGGTGCCG	CAAAGCTGG	TGGAGTGC	TCTTCCTGAG	GGCGATACGG	TCGTCGTCCC	6840
6841	CTCAAACCTGG	CAGATGCACG	GTACGATGC	GCCCCATCTAC	ACCAACGTAA	CCTATCCCCT	6900
6901	TACGGTCAAT	CCGCCGTTTG	TTCCCACCGGA	GAATCCGACG	GGTTGTTACT	CGCTCACATT	6960
6961	TAATGTTGAT	GAAAGCTGG	TACAGGAAGG	CCAGACGCGA	ATTATTTTG	ATGGCGTTCC	7020
7021	TATTGGTTAA	AAAATGAGCT	GATTTAACAA	AAATTTAACG	CGAATTTAA	CAAATATTA	7080
7081	ACGTTACAA	TTTAAATATT	TGCTTATACA	ATCTTCTCT	TTTGGGGT	TTTCTGATTA	7140
7141	TCAACCGGGG	TACATATGAT	TGACATGCTA	GTTTACGAT	TACCGTTCAT	CGATTCTCTT	7200
7201	GTTCGCTCCA	GAATCTCAGG	CAATGACCTG	ATAGCCTT	AGATCTCTC	AAAAATAGCT	7260
7261	ACCCCTCTCCG	GCATTAATT	ATCAGCTAGA	ACGGTTGAAT	ATCATATTGA	TGGTATTG	7320
7321	ACTGTCTCCG	GCCTTTCTCA	CCCTTTGAA	TCTTACCTA	CACATTACTC	AGGCATGCA	7380
7381	TTTAAATAT	ATGAGGGTTC	TAAAAATT	TATCCTGCG	TGAAATAA	GGCTTCTCCC	7440
7441	GCAAAGTAT	TACAGGGTCA	TAATGTTTT	GGTACAACCG	ATTAGCTT	ATGCTCTGAG	7500
7501	GCTTATTG	TTAATTG	TAATTCTTTG	CCTTGCGCTGT	ATGATT	GGATGTT	7557
	10	20	30	40	50	60	

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC	AAATGAAAAT
61	ATAGCTAAC	AGGTTATTGA	CCATTGCGGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT
121	CGTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTA
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAAGATT	AGCAATTAAAG	CTCTAAGCCA
241	TCTGCAAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG
301	TTGGAGTTTG	CTTCCGGTCT	TCTTTTGAT	GAAGCTCGAA	TTAAAACGCG	ATATTGAAG
361	TCTTCGGGC	TTCCCTCTAA	TCTTTTGAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT
421	CAGGGTAAAG	ACCTGATTT	TGATTTATGG	TCATTCTCGT	TTCTGAACT	GTTTAAACGAA
481	TTTGAGGGGG	ATTCAATGAA	TATTATGAC	GATTCCGCG	TATTGGACGC	TATCCAGTCT
541	AAACATTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTG	CAAAGGCCTC	TCGCTATTT
601	GGTTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCTCTCGT
661	AATTCCCTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCTAA	ATCTCAACTG
721	ATGAATCTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATTAA
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGT
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC
1021	TGTACACCCT	TCATCTGTCC	TCTTCAAGG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC
1081	GTCTGCGCCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTGCG	CGGAGTTTCA	CACAATTAT
1141	CAGGGATGA	TACAAATCTC	CGTTGACTT	TGTTTCCGCG	TTGGTATAAT	CGCTGGGGGT
1201	CAAAGATGAG	TGTTTAGTG	TATTCTTTCG	CCTCTTCTG	TTAGGTTGG	TGCCCTCGTA
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACCTCCTC	ATGAAAAAGT	CTTCTAGTCT
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA
1381	CGATCCCGCA	AAAGCGGCCT	TTAACCTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA
1441	TGCGTGGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA
1501	ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTTT
1561	TTTTGGAGA	TTTCAACGT	GAAAAAAATTAA	TTATTGCAA	TTCTCTTGT	TGTTCCCTTC
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAATC	TTAGATCGT	ACGCTAACTA	TGAGGGTTGT
1741	CTGTGGAATG	CTACAGGCGT	TGTTAGTTGT	ACTGGTGACG	AAACTCAGT	TTACGGTACA
1801	TGGGTTCTA	TGGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT
1921	ATTCCGGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACTT	ATCCGCTCTGG	TACTGAGCAA
1981	AAACCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAAATAC	TTTCATGTTT
2041	CAGAATAATA	GGTCCGAAA	TAGGCAGGGG	GCATTAACCTG	TTTATACGGG	CACTGTTACT
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG
2161	TATGACGCTT	ACTGGAAACGG	TAAATTCTAGA	GAUTGCCTT	TCCATTCTGG	CTTTAATGAA
2221	GATCCATTCTG	TTTGTGATAA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT
2281	GCTGGCGGCG	GCTCTGGTGG	TGGTTCTGGT	GGCAGGCTCTG	AGGGTGGTGG	CTCTGAGGGT
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCAGGTTCCG	GTGGTGGCTC	TGGTTCCGGT
2401	GATTTTGATT	ATGAAAAGAT	GGCAAACGCT	ATAAGGGGG	CTATGACCGA	AAATGCCGAT
2461	GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTGCTAC	TGATTACGGT
2521	GCTGCTATCG	ATGGTTTCAT	TGTTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT
2581	GGTGATTTG	CTGGCTCTAA	TTCCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCAACCT
2641	TTAATGAATA	ATTTCGCTCA	ATATTACCT	CCCCCTCC	AATCGGTTGA	ATGTCGCCCT
2701	TTTGTCTTTA	GCGCTGGTAA	ACCATATGAA	TTCTCTTATG	ATTGTGACAA	AATAAACTTA
2761	TTCCGTGGTG	TCTTTCGCTT	TCTTTTATAT	GTTGCCACCT	TTATGTATGT	ATTTTCTACG
2821	TTTGCTAACAA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCGT
2881	TATTATTGCG	TTTCTCTGGT	TTCTCTCTGG	TAACTTTGTT	CGGCTATCTG	CTTACTTTT
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTCATT	GTTTCTTGCT	CTTATTATTG
3001	GGCTTAACTC	AATTCTTGTG	GGTTATCTCT	CTGATAATTAG	CGCTCAATTAA	CCCTCTGACT
3061	TTGTCAGGG	TGTTCACTTA	ATTCTCCCGT	CTAATGCGCT	TCCCTGTTTT	TATGTTATT
3121	TCTCTGAAA	GGCTGCTATT	TTCATTTTG	ACGTTAAACA	AAAAATCGTT	TCTTATTATTG
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTGTT	GTAACTGGCA	AAAATCGTT	TCTTATTATTG
3241	CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT
3301	CTTGATTAA	GGCTTCAAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAAC	GCCTCGCGTT
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTG	CTATTGGGCG	CGGTAATGAT
3421	TCCTACGATG	AAAATAAAAAA	CGGCTTGCCT	GTCTCGATG	AGTGGGGTAC	TGGGTTTAAT
3481	ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CGGATTATTG	ATTGGTTTCT	ACATGCTCGT
3541	AAATTAGGAT	GGGATATTAT	TTTCTCTGGT	CAGGACTTAT	CTATTGTTG	TAAACAGGGCG
3601	CGTTCTGCT	TAGCTGAACA	TGTTGTTTAT	TGTCGTCGTC	TGGACAGAAAT	TACTTTACCT
3661	TTTGTGCTA	CTTATATTCT	TCTTATTACT	GGCTCGAAAAA	TGCCTCTGCC	TAAATTACAT
3721	GGTGGCGTTG	TAAATATGG	CGATTCTCAA	TAAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT
3781	ACTGGTAAGA	ATTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT
3841	CCGGTGT	ATTCTTATT	AACGCCCTAT	TTATCACACG	GTCGGTATT	CAAACCAATTAA
3901	TTTGTGCTA	GAAGATGAA	GCTTACTAA	ATATATTTGAA	CGGCGT	CGGCGT

11 / 11

4081 CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT 4140
 4141 AGCGACGATT TACAGAAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTCC 4200
 4201 ATTAaaaaaaAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTTTGTT TCTTGATGTT 4260
 4261 TGTTTCATCA TCTTCTTTG CTCAGGTAAT TGAAATGAAT AATTGCGCTC TGCGCGATT 4320
 4321 TGTAACCTGG TATTCAAAGC AATCAGGCAG ATCCGTTATT GTTTCTCCCG ATGTAAAAGG 4380
 4381 TACTGTTACT GTATATTCACT CTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTTATTTC 4440
 4441 TGTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCATAA TTCAGAAGTA 4500
 4501 TAATCCAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA 4560
 4561 TGATAATTCC GCTCCTCTG GTGGTTCTT TGTTCCGCAA AATGATAATG TTACTCAAAC 4620
 4621 TTTTAAAATT AATAACGTT GGGCAAAGGA TTTAATACGA GTTGTGAAAT TGTTTGAAA 4680
 4681 GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAAC TATTAGTTGT 4740
 4741 TAGTGCACCT AAAGATATT TAGATAACCT TCCTCAATT CTTCTACTG TTGATTGCCC 4800
 4801 AACTGACCAAG ATATTGATG AGGGTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTAA 4860
 4861 TTTTCACTT GCTGCTGGT CTCAGCGTGG CACTGTTGCA GGCCTGTTA ATACTGACCG 4920
 4921 CCTCACCTCT GTTTTATCCT CTGCTGGTGG TTCGTTCGGT ATTTTAATG GCGATGTTTT 4980
 4981 AGGGCTATCA GTTGCACAT TAAAGACTAA TAGCCATCA AAAATATTGT CTGTGCCACG 5040
 5041 TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTAT 5100
 5101 TACTGGTCGT GTGACTGGT AATCTGCCAA TGAAATAAT CCATTTCAAGA CGATTGAGCG 5160
 5161 TCAAAATGTA GGTATTTCGA TGAGCTTTT TCCCTGTTGCA ATGGCTGGCG GTAATATTGT 5220
 5221 TCTGGATATT ACCAGCAAGG CGATAGTTT GAGTTCTCT ACTCAGGCAA GTGATGTTAT 5280
 5281 TACTAATCAA AGAAGTATTG CTACAAACGGT TAATTGCGT GATGGACAGA CTCTTTACT 5340
 5341 CGGTGGCTC ACTGATTATA AAAACACTTC TCAAGATCTT GGCCTACCGT TCCCTGCTAA 5400
 5401 AATCCCCTTA ATCGGCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT 5460
 5461 ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCGCATT AGCGCAGCG 5520
 5521 GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTGCGAG CGCCCTAGCG CCCGCTCCTT 5580
 5581 TCGCTTCTT CCCTCCTT CTCGCCACGT TGCCGGCTT TCCCGTCAA GCTCTAAATC 5640
 5641 GGGGGCTCCC TTAGGGTTC CGATTAGTG CTTTACGGCA CCTCGACCCC AAAAAACTTG 5700
 5701 ATTTGGGTGA TGTTTACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTGA 5760
 5761 CGTTGGAGTC CACGTTCTT AATAGTGGAC TCTTGTCCA AACTGGAACA ACACTCAACC 5820
 5821 CTATCTGGG CTATTCTTT GATTATAAG GGATTGGTCC GATTTCGGAA CCACCATCAA 5880
 5881 ACAGGATTTT CGCCTGCTGG GGCAAACCAAG CGTGGACCGC TTGCTGCAAC TCTCTCAGGG 5940
 5941 CCAGGCGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCAACCT 6000
 6001 GGCGCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TGCACTGGC 6060
 6061 ACGACAGGTT TCCCAGTGG AAAGGGGCA GTGAGCGCAA CGCAATTAAAT GTGAGTTAGC 6120
 6121 TCACTCATTA GGCACCCCCAG GCTTACACT TTATGCTTCC GGCTCGTATG TTGTGTTGAA 6180
 6181 TTGTGAGCGG ATAACAATTG CACAGCCAA GGAGACAGTC ATAATGAAAT ACCTATTGCC 6240
 6241 TACGGCAGCC GCTGGATTGT TATTACTCGC TGCCCAACCA GCGATGGCG AGCTCTCCC 6300
 6301 GCCATCTGAT GAGCAGTGTG AATCTGGAAC TGCTCTGTT GTGTGCTGC TGAATAACTT 6360
 6361 CTATCCCAGA GAGGCCAAAG TACAGTGGAA GGTGGATAAC GCCCTCCAAT CGGGTAACTC 6420
 6421 CCAGGAGAGT GTCACAGAGC AGGACAGCAA GGACAGCACC TACAGCCTCA GCAGCACCT 6480
 6481 GACGCTGAGC AAAGCAGACT ACGAGAAACA CAAAGTCTAC GCCTGCGAAG TCACCCATCA 6540
 6541 GGGCCTGAGC TCGCCCGTCA CAAAGAGCTT CAAACAGGGGA GAGTGTCTA GAACGCGTCA 6600
 6601 CTTGGCACTG GCCGTCGTTT TACAACGTC TGACTGGAA AACCCCTGGCG TTACCCAAAGC 6660
 6661 TTTGTACATG GAGAAAATAA AGTGAACAA AGCACTATTG CACTGGCACT CTTACCGTTA 6720
 6721 CTGTTTACCC CTGTGGCAAA AGCCGCTCC ACCAAGGGGC CATCGGTCTT CCCCCCTGGCA 6780
 6781 CCCTCCTCCA AGAGCACCTC TGGGGCACA GCGGCCCTGG GCTGCTGGT CAAGACTAAT 6840
 6841 TCCCCGAACC GGTGACGGTG TCGTGGAACT CAGGGCCCT GACCAAGCGC GTGCACACCT 6900
 6901 TCCCGGCTGT CCTACAGTCC TCAGGACTCT ACTCCCTCAG CAGCGTGGTG ACCGTGCCCT 6960
 6961 CCAGCAGCTT GGGCACCCAG ACCTACATCT GCAACGTGAA TCACAAGGCC AGCAACACCA 7020
 7021 AGGTGGACAA GAAAGCAGAG CCAAATCTT GTACTAGTGG ATCCCTACCG TACGACGTT 7080
 7081 CGGACTACGC TTCTTAGGCT GAAGGGATG ACCCTGCTAA GGCTGCATT AATAGTTTAC 7140
 7141 AGGCAAGTGC TACTGAGTAC ATTGGCTACG CTTGGCTAT GGTAGTAGTT ATAGTTGGTG 7200
 7201 CTACCATAGG GATTAATTG TTCAAAAAGT TTACGAGCAA GGCTCTTAA GCAATAGCGA 7260
 7261 AGAGGCCCGC ACCGATCGCC CTTCCCAACA GTTGCACAGC CTGAATGGCG AATGGCGCTT 7320
 7321 TGCCCTGGTT CGGGCACCAAG AAGCGGTGC GGAAAGCTGG CTGGAGTGC ATCTTCTGA 7380
 7381 GGCCGATACG GTCGTCGTC CCTCAAACGT GCAGATGCAC GTTACGATG CGCCCATCTA 7440
 7441 CACCAACGTA ACCTATCCCA TTACGGTCAA TCCGCCGTGTT GTTCCCACGG AGAATCCGAC 7500
 7501 GGGTTGTAC TCGCTCACAT TTAATGTTGA TGAAAGCTGG CTACAGGAAG GCGAGACGCG 7560
 7561 AATTATTTT GATGGCGTTC CTATTGGTTA AAAAATGAGC TGATTAACA AAAATTTAAC 7620
 7621 GCGAATTAA ACAAAATATT AACGTTACA ATTAAATAT TTGCTTATAC ATCTTCTG 7680
 7681 TTTGGGGC TTTCTGATT ATCAACGGG GTACATATGA TTGACATGCT AGTTTACGA 7740
 7741 TTACCGTCTCA TCGATTCTCT TGTTGCTCC AGACTCTCAG GCAATGACCT GATGCCCC 7800
 7801 GTAGATCTCT CAAAAATAGC TACCCCTCTC GGCAATTAA TATCAGCTAG AACGGTTGAA 7860
 7861 TATCATATTG ATGGTATTG TACTGCTCC GGCCCTTCTC ACCCTTTGA ATCTTACCT 7920
 7921 ACACATTACT CAGGCATTGC ATTAAAATA TATGAGGGT CTTAAAATTT TTATCCTGC 7980
 7981 GTTGAATTA AGGCCTCTC CGCAAAGT TTACAGGTT TTTGTTT TGGTACAC 8040

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US91/07140

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all)

According to International Patent Classification (IPC) or to both National Classification and IPC
 IPC(5): C12N 15/64, 15/70
 U.S.CI.: 435/252.3, 320.1

II. FIELDS SEARCHED

Classification System	Minimum Documentation Searched	
		Classification Symbols
U.S.CI.		435/69.7, 172.3, 252.3, 320.1

Documentation Searched other than Minimum Documentation
 to the Extent that such Documents are Included in the Fields Searched

APS, STN/MEDLINE. TERMS USED: SURFACE EXPRESSION VECTOR#, DIRECTED EVOLUTION. SINGLE CHAIN ANTIBOD?.

III. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages	Relevant to Claim No.
Y	W.C. A. 08/00800 (FOM ET AL) 07 September 1988, 1-75 see entire document.	
Y	Nucleic Acids Research, Vol. 12, No. 9, issued SEPTEMBER 1984, BOSS ET AL, "Assembly of functional antibodies from immunoglobulin heavy and light chains synthesized in <u>E. coli</u> ", pages 3731-3806, see the abstract.	5-75
Y	Proceedings of the National Academy of Sciences, Vol. 86, issued AUGUST 1989, SASTRY ET AL, "Cloning of the immunological repertoire in <u>Escherichia coli</u> for generation of monoclonal catalytic antibodies: Construction of a heavy chain variable-region specific cDNA library", pages 5720-5722, see the abstract.	1-75
Y	Science, Vol 246, issued 08 December 1989, Huse et al, "Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire in Phage Lambda", pages 1275-1281, see entire document.	1-75

* Special categories of cited documents: ¹⁰

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "C" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"Z" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search

06 January 1992

Date of Mailing of this International Search Report

21 JAN 1992

International Searching Authority

ISA US

Signature of Authorized Officer

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Gene, Vol. 70, issued 1998, PARMLEY ET AL.
"Antibody-selectable filamentous fd phage
vectors: affinity purification of target
genes", pages 105-118, see entire document.

6-75

V. OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. Claim numbers _____ because they relate to subject matter not required to be searched by this Authority, namely:

2. Claim numbers _____ because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claim numbers _____ because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This International Searching Authority found multiple inventions in this international application as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

The additional search fees were accompanied by applicant's protest.
 No protest accompanied the application.